

Eine anschauliche Methode zur mehrdimensionalen Ausreißererkennung in der Kieferorthopädie

Diplomarbeit am Institut für Mathematik, Schwerpunkt Stochastik
Fachbereich Informatik und Mathematik
J. W. Goethe Universität Frankfurt a. M.

von

Markus Bingmer

Betreuerin:

JProf. Dr. Gaby Schneider

4. Juni 2008

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich bei dieser Arbeit so tatkräftig unterstützt haben.

Frau JProf. Gaby Schneider danke ich für die intensive Betreuung, dass ich jederzeit an ihre Tür klopfen durfte und dass sie bei Fragen immer für mich da war.

Herrn Dr. Brooks Ferebee danke ich für das interessante Thema dieser Diplomarbeit und für seine unterstützende Beratungstätigkeit. Von ihm kam die Idee, die hinter der hier vorgestellten MH Methode steckt.

Bei Herr Dr. Volkan Özkan und Herr Dr. Yong-min Jo bedanke ich mich für die gute Zusammenarbeit, für das Bereitstellen der Daten sowie für die Vermittlung des nötigen kieferorthopädischen Hintergrundes.

Nicht zuletzt gilt mein besonderer Dank meinen Eltern, die mir das Studium ermöglicht haben und mich in dieser langen Zeit immer unterstützten.

Zusammenfassung

Kieferorthopäden beschreiben die Anordnung der Zähne und die Stellung der Kiefer üblicherweise mittels Winkel und Strecken in der sagittalen Gesichtsebene. Im vorliegenden Fall werden fünf Winkel $\mathbf{X}_1, \dots, \mathbf{X}_5$ betrachtet und jedes Individuum \mathbf{x} lässt sich als Punkt in einem 5-dimensionalen Raum darstellen. Individuen, die laut Experten ein gut funktionierendes Gebiss und ein harmonisches Äußeres besitzen, formen eine Punktwolke, die im Folgenden als die Norm Population bezeichnet wird. Individuen fern von der Wolke benötigen kieferorthopädische Behandlung.

Welche Form sollte dieser Eingriff annehmen? Durch Hilfsmittel der modernen Kieferorthopädie lassen sich die beschriebenen Winkel nahezu nach Belieben ändern. Dies ist natürlich verbunden mit einer unterschiedlichen Menge an Problemen, Arbeitsaufwand und Unannehmlichkeiten, abhängig vom individuellen Patienten.

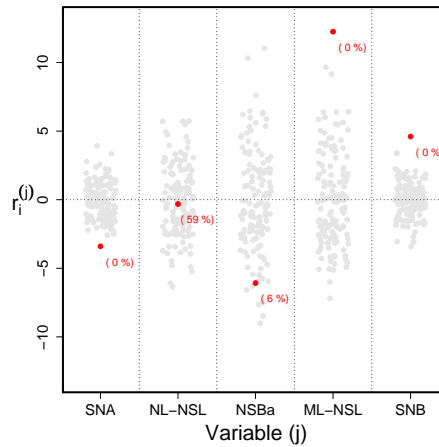
Diese Arbeit präsentiert eine Methode, die auf jedem Computer leicht implementierbar und auf k Variablen verallgemeinerbar ist. Sie ermöglicht Kieferorthopäden eine Visualisierung, wie verschiedene denkbare Anpassungen der Winkel eines Patienten dessen relative Position zur Norm Population verändern. Damit unterstützt sie Kieferorthopäden bei der Entscheidung für einen Behandlungsplan, der die besten Ergebnisse verspricht.

Das **erste** Kapitel widmet sich einer kurzen Einführung in die Kieferorthopädie und die oben geschilderte Problemstellung. Die aufgeführten kieferorthopädischen Fragen werden in einen mathematischen Kontext übertragen und an einem einführenden Beispiel erläutert.

Der Versuch, die Fragen zu beantworten, erfolgt im **zweiten** Kapitel. Es werden einige grundlegende Begriffe der linearen Regression eingeführt sowie der *Multi-Harmonie-Ansatz* oder kurz *MH Ansatz* in einem allgemeinen Rahmen für k Variablen vorgestellt. Dieser basiert auf der Betrachtung der folgenden multivariaten linearen Regressionsgleichungen:

$$\begin{array}{rcl}
\mathbf{X}_1 & = & f_1(\mathbf{X}_2, \dots, \mathbf{X}_k) \quad + \mathbf{F}^{(1)} \\
\vdots & & \vdots \\
\mathbf{X}_j & = & f_j(\mathbf{X}_1, \dots, \mathbf{X}_{j-1}, \mathbf{X}_{j+1}, \dots, \mathbf{X}_k) \quad + \mathbf{F}^{(j)} \\
\vdots & & \vdots \\
\mathbf{X}_k & = & f_k(\mathbf{X}_1, \dots, \mathbf{X}_{k-1}) \quad + \mathbf{F}^{(k)}
\end{array}$$

Um harmonische Winkelkombinationen zu quantifizieren, werden die Parameter mittels n Individuen der Norm Population geschätzt. Von zentraler Bedeutung sind die Residuen r_i der Regressionen. Diese ermöglichen ein zu betrachtendes Individuum \mathbf{x} mit den Normalen zu vergleichen und anhand einer Regel λ zu klassifizieren. Um dies zu visualisieren, wird eine sogenannte Abweichungsgrafik eingeführt, welche eine stark abgewandelte Form eines *Parallelen Koordinaten Plots* [ID90] darstellt:



Die Methode wird im weiteren Verlauf nochmals ausführlich für $k = 2$, anhand von Grafiken und eines Beispiels erläutert und zu guter Letzt auf den koreanischen Datensatz, der in Anhang A näher beschrieben wird, angewendet.

Eine anschauliche geometrische Interpretation der allgemeinen MH Methode im Raum der Variablen erfolgt im **dritten** Kapitel. Da man sich im allgemeinen Fall in einem k -dimensionalen Variablenraum befindet, stellen die oben genannten Gleichungen $(k - 1)$ -dimensionale Hyperebenen dar. Es werden die möglichen beobachtbaren Fälle der Patientenklassifikation in der Abweichungsgrafik am Beispiel $k = 2$ aufgezeigt und diskutiert. Des Weiteren werden das Verhalten des MH Ansatzes beim Ändern der Winkelwerte und die Auswirkungen auf die anderen Variablen kurz erläutert.

Im **vierten** Kapitel wird eine aus der Kieferorthopädie bekannte Methode vorgestellt - die *Harmoniebox* nach Segner [SA98]. Es erfolgt eine kurze Einführung in die Konstruktion der Box und ihre Anwendung. Darüber hinaus wird eine geometrische Interpretation geliefert sowie einige Probleme aufgezeigt. Die Harmoniebox basiert auf einfachen linearen Regressionsgleichungen der Form

$$\begin{aligned} \mathbf{X}_2 &= \alpha_0^{(2)} + \alpha_1^{(2)} \mathbf{X}_1 + F^{(2)} \\ &\vdots \\ \mathbf{X}_5 &= \alpha_0^{(5)} + \alpha_1^{(5)} \mathbf{X}_1 + F^{(5)}. \end{aligned}$$

Für vorgegebene Werte von \mathbf{X}_1 werden die dazugehörigen Vorhersagen bestimmt und tabellarisch zusammengefasst. Harmonische Winkelkombinationen stellen nach Segner alle Werte auf einer horizontalen Linie in dieser Tabelle dar. Für die Klassifikation eines Individuums wird die Linie mittels einer bestimmten Vorgehensweise positioniert, wobei Abweichungen von jeweils einem Standardfehler der entsprechenden multivariaten Regression, wie sie oben beschrieben wurden, zugelassen werden. Das Ganze wird grafisch visualisiert:

ANB	SNA	NL-NSL	NSBa	ML-NSL	SNB	ML-NL
2	60	24			58	27
	62	23	150	49	60	
	64	22	148	47	62	26
	66	21			64	
	68	20	146	45	66	25
	70	19	144	43	68	
	72	18	142	41	70	24
	74	17	140	39	72	
	76	16	138	37	74	23
	78	15	136	35	76	
	80	14	134	33	78	22
	82	13	132	31	80	
	84	12	130	29	82	21
	86	11	128	27	84	
	88	10	126	25	86	20
2.5	90	9	124	23	88	19
	92	8	122	21	90	
	94	7	120	19	92	18
	96	6	118	17	94	
	98	5			96	
	100	4				

Das **fünfte** Kapitel beschäftigt sich mit dem Vergleich des MH Ansatzes und der *Harmoniebox*. Im weiteren Verlauf werden die beiden Ansätze in Bezug auf die Regressionsgleichungen und die Klassifikationsgüte beim koreanischen Datensatz im Sinne der Sensitivität und Spezifität einander gegenübergestellt. Es zeigt sich, dass der MH Ansatz besser als die Harmoniebox dazu geeignet ist Patienten und Normale aus dem Datensatz zu erkennen.

Schließlich wird ein wichtiger Zusammenhang zwischen den hier gezeigten bivariaten und multivariaten Gleichungen aufgedeckt. Harmonische Winkelkombinationen der Harmoniebox sind bis auf einige Ausnahmen auch im MH Ansatz harmonisch. Dies wird im letzten Abschnitt des Kapitels formal in einem allgemeinen Satz gezeigt.

Abgerundet wird die Arbeit im **sechsten** und letzten Kapitel mit einer kurzen Diskussion der hier vorgestellten Methode. Es werden keine Handlungsempfehlungen für Kieferorthopäden gegeben, sondern eine Darstellung der momentanen Situation des Individuums, was der weiteren Behandlungsplanung dient.

Inhaltsverzeichnis

1. Problemstellung	1
1.1. Patientenbewertung in der Kieferorthopädie	1
1.2. Mathematischer Kontext	3
2. Der Multi-Harmonie-Ansatz	7
2.1. Grundlegende Begriffe von Regressionsmodellen	8
2.2. Der allgemeine Fall	9
2.2.1. Vorgehensweise	10
2.2.2. Klassifikationsregeln	11
2.3. Der Ansatz für $k=2$	14
2.4. Anwendung auf den Datensatz	18
2.4.1. Regressionen	18
2.4.2. Visualisierung der Residuen	20
2.4.3. Visualisierung von Anpassungen	22
3. Interpretation und Eigenschaften	25
3.1. Geometrische Betrachtung	25
3.2. Möglichkeiten der Klassifikation	29
3.3. Änderung der Abweichungen	31
4. Die Harmoniebox nach Segner	35
4.1. Konstruktion der Box	35
4.2. Das Harmonieschema	38
4.3. Eigenschaften	41
4.3.1. Probleme der Harmoniebox	41
4.3.2. Geometrische Interpretation	43
5. Vergleich von Harmoniebox und MH	47
5.1. Vergleich	47
5.1.1. Regressionen	48
5.1.2. Klassifikation von Individuen	48
5.2. Zusammenhang zwischen den Ansätzen	50
5.2.1. Harmonielinien im MH Ansatz	50

5.2.2. Mathematischer Zusammenhang	52
6. Diskussion	55
A. Der Datensatz	59
A.1. Die Variablen	59
A.2. Deskriptive Statistiken	62
A.2.1. Verteilungen	62
A.2.2. Korrelationen in der Normgruppe	62
Literaturverzeichnis	67
Abbildungsverzeichnis	69
Tabellenverzeichnis	70

Kapitel 1.

Problemstellung

Dieses Kapitel widmet sich einer kurzen Einführung in die Kieferorthopädie und einer dort vorkommenden Problemstellung in Bezug auf die Bewertung und Diagnose von Patienten mit Hilfe der Kephalometrie. Die im ersten Abschnitt aufgeführten kieferorthopädischen Fragen werden im zweiten in einen mathematischen Kontext übertragen und an einem einführenden Beispiel erläutert.

1.1. Patientenbewertung in der Kieferorthopädie

Die Kieferorthopädie ist ein Teilgebiet der Zahnmedizin, das sich mit der Prophylaxe, Erkennung und Behandlung von Fehlstellungen der Kiefer und der Zähne befasst. Zu ihren Aufgaben gehört neben der Sicherstellung eines funktionierenden Kau- und Sprachmechanismus mittlerweile auch das Herstellen von Ästhetik und Harmonie im Gesicht. Dieser Auftrag wird durch moderne Techniken (z.B. Spangen oder Operationen) verwirklicht, mit deren Hilfe sich die relativen Positionen der Zähne und Kiefer nahezu nach Belieben ändern lassen. Natürlich ist dies verbunden mit einer unterschiedlichen Menge an Problemen, Arbeitsaufwand und Unannehmlichkeiten, abhängig vom individuellen Patienten.

Eines dieser wichtigen Hilfsmittel zur Diagnose des Zustandes eines Individuums stellt das Fernröntgenseitenbild dar. Aufgenommen, um Abbildungsfehler zu minimieren, aus einer für Praxen üblichen Entfernung von mind. $1.7m$, kann der Kieferorthopäde Punkte im Schädel identifizieren sowie Winkel und Strecken in der sagittalen (seitlichen) Gesichtsebene einer Person messen. Die Kopfvermessungskunde, auch *Kephalometrie*, kennt Hunderte solcher Variablen.

Dieser Arbeit liegen die kephalometrischen Daten von 220 koreanischen Individuen zugrunde, bei denen die Winkel SNA , $NL-NSL$, $NSBa$, $ML-NSL$ und SNB bestimmt wurden

(für einen kurzen Überblick sei auf Anhang A verwiesen). Dieser Datensatz teilt sich in folgende zwei Gruppen auf:

„**Normale**“: 134 Individuen, die von Experten als einen guten Biss und ästhetisches Äußeres habend, klassifiziert wurden.

„**Patienten**“: 86 Individuen, die nach Ansicht der Beobachter behandlungsbedürftig sind.

Der Kieferorthopäde muss sich bei einem Individuum die folgenden Fragen stellen:

1. Ist das Individuum normal oder behandlungsbedürftig?
2. Gegeben, es handelt sich bei der Person um einen Patienten - warum ist es einer? Welche kephalometrischen Variablen sind dafür verantwortlich?
3. Wie könnte man das Individuum mit Zielsetzung einer harmonischen Winkelkombination behandeln? Welche Möglichkeiten hat der Kieferorthopäde?

Vorherige Methoden [Ste53] benutzten Populationsmittelwerte von Normalen als Referenz für die Analyse der kephalometrischen Daten. Autoren wie Solow [Sol66] haben gezeigt, dass diese Vorgehensweise nicht mehr zeitgemäß ist, da die offensichtlich bestehenden Korrelationen und Zusammenhänge zwischen den Winkeln nicht berücksichtigt werden.

In späteren Arbeiten gewinnen die sogenannten *floating norms*, die ihren Durchbruch durch die Arbeit von Segner [Seg89] und seine in Kapitel 4.1 vorgestellte Harmoniebox hatten, zunehmend an Bedeutung. Sie stellen individuelle Normen unter Berücksichtigung des jeweiligen Gesichtstyps des Patienten dar und basieren weitestgehend auf einfachen linearen Regressionsmodellen. Doch es wird sich zeigen, dass auch diese wichtige mehrdimensionale Zusammenhänge vernachlässigen und aufgrund ihrer Vielzahl an möglichen Modifikationen problematisch sind.

Aus diesen Gründen wird in dieser Arbeit zu multivariaten Regressionsmodellen übergegangen und ein Ansatz vorgestellt, mit dem versucht wird, die Verhältnisse, in denen sich die Winkel der Normalen befinden, zu quantifizieren. Das Ziel der hier vorgestellten Methode ist es, die oben genannten Fragen anhand einer Referenzgruppe (in unserem Fall die Normalen) zu beantworten. Für den Anwender soll der Ansatz leicht interpretierbare Ergebnisse liefern, grafisch visualisierbar und auf einem Computer implementierbar sein und die Kieferorthopäden somit in der Wahl des besten Behandlungsplanes unterstützen.

1.2. Mathematischer Kontext

Es sei $N := \{1, \dots, n\}$ die Menge der vorliegenden Normalen (die Referenzgruppe) und k die Anzahl der vorliegenden Winkel. Man hat somit Beobachtungen x_{ij} mit $i = 1, \dots, n$ und $j = 1, \dots, k$, die sich in einer $n \times k$ Datenmatrix \mathbf{X} zusammenfassen lassen:

$$\mathbf{X} := \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nk} \end{pmatrix}.$$

Die i -te Zeile repräsentiert das i -te normale Individuum und die j -te Spalte den j -ten Winkel. Bezieht man sich auf den j -ten Winkel sämtlicher vorliegenden Normalen, so wird er zu folgendem Spaltenvektor zusammengefasst:

$$\mathbf{X}_j := \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix}.$$

Bezieht sich der Text auf die Werte des i -ten Normalen, so wird $\mathbf{x}_i := (x_{i1}, \dots, x_{ik})$ geschrieben. Ein nicht klassifiziertes Individuum (nicht bekannt, ob normal oder behandlungsbedürftig) wird dargestellt als $\mathbf{x} := (x_1, \dots, x_k)$ (ohne Index).

Angenommen, man habe für jedes Individuum $i = 1, \dots, n$ die Winkel x_{i1}, \dots, x_{ik} gegeben, dann lässt sich jedes Individuum als Punkt im \mathbb{R}^k auffassen. Die Referenzgruppe bildet somit eine Punktwolke mit Schwerpunkt $\bar{\mathbf{x}} := (\bar{x}_1, \dots, \bar{x}_k)$ in diesem Raum. Dabei bezeichnet $\bar{x}_j := \frac{1}{n} \sum_{i=1}^n x_{ij}$ den (Stichproben-)Mittelwert des j -ten Winkels ($j = 1, \dots, k$).

Die Fragen aus Abschnitt 1.1 lassen sich wie folgt übernehmen:

1. Ist ein zusätzlicher Punkt $\mathbf{x} \in \mathbb{R}^k$ innerhalb der Punktwolke?
2. Falls nein - warum nicht?
3. Welche Möglichkeiten gibt es, in die Normwolke zu gelangen?

Geht man davon aus, dass die Winkel innerhalb der Norm Population normalverteilt sind, so kann man sie als \mathbb{R} -wertige Zufallsvariablen X_{i1}, \dots, X_{ik} auffassen. Es bietet sich

dann sogar an $(X_{i1}, \dots, X_{ik})^T \sim \mathcal{N}(\mu, \Sigma) \forall i$ anzunehmen, wobei $\mathcal{N}(\mu, \Sigma)$ die mehrdimensionale Normalverteilung mit Erwartungswert $\mu \in \mathbb{R}^k$ und Kovarianzmatrix Σ bezeichnet. Des Weiteren kann man davon ausgehen, dass die Zufallsvektoren $(X_{11}, \dots, X_{1k})^T, (X_{21}, \dots, X_{2k})^T, \dots, (X_{n1}, \dots, X_{nk})^T$ unabhängig sind.

Unter diesen Annahmen ist es naheliegend, die Mahalanobis Distanz d_M als Metrik im \mathbb{R}^k zu verwenden, welche für zwei Vektoren $x, y \in \mathbb{R}^k$ wie folgt definiert ist:

$$d_M(x, y) := \sqrt{(x - y)^T \Sigma^{-1} (x - y)}. \quad (1.1)$$

Die Mahalanobis Distanz ist gut dazu geeignet, die erste Frage zu beantworten. Eine Möglichkeit dies umzusetzen wäre z.B. zu überprüfen, ob für den Punkt $\mathbf{x} \in \mathbb{R}^k$ gilt:

$$d_M(\mathbf{x}, \bar{\mathbf{x}}) \leq \max_{i \in N} d_M(\mathbf{x}_i, \bar{\mathbf{x}}).$$

In diesem Fall wäre der Abstand von \mathbf{x} zum Schwerpunkt kleiner als der Abstand des am weitesten entfernten Normalen, was darauf schließen lassen könnte, dass auch \mathbf{x} innerhalb der Punktwolke liegt. Doch leider scheitert die Mahalanobis Distanz an den anderen beiden Fragen, weil man keine Rückschlüsse ziehen kann, warum ein Punkt außerhalb der Wolke liegt. Sie wird dennoch im Weiteren als Referenz zum Vergleich herangezogen.

Eine weitere denkbare Vorgehenseise, um ein neues Individuum zu klassifizieren und die oben gestellten Fragen zu beantworten, wäre die Verteilung der einzelnen Variablen innerhalb der Referenzgruppe zu untersuchen und die Werte des zusätzlichen Objekts eindimensional mit dem jeweiligen Mittelwert \bar{x}_j zu vergleichen. Dies entspricht dem Vergleich mit Populationsmittelwerten und vernachlässigt, wie bereits erwähnt, die Korrelationen zwischen den Variablen.

Beispiel 1.1. *In Abbildung 1.1 sind die Verteilungen von zwei fiktiven Variablen der Referenzgruppe anhand von Histogrammen dargestellt, wobei die Werte eines zusätzlichen Individuums markiert sind. Da man keine größeren Abweichungen vom jeweiligen Mittelwert feststellen kann, wird es als in der Normgruppe liegend angenommen. Der Sachverhalt wird in Abbildung 1.1c nochmals mittels eines Streudiagramms betrachtet. Das Individuum liegt deutlich außerhalb der durch die graue Ellipse dargestellten Referenzgruppe. ■*

Dieses Problem und die Fragestellungen sind in zwei Dimensionen noch relativ leicht anschaulich mittels eines Streudiagramms darstellbar, werden mit mehr Variablen jedoch

zunehmends komplexer und benötigen eine für Kieferorthopäden zugeschnittene Methode.

Eine mögliche Vorgehensweise wird von Segner [Seg89] in Form seiner Harmoniebox vorgestellt, die in Kapitel 4.1 ausführlich betrachtet wird. Sein Ansatz berücksichtigt Korrelationen dadurch, dass alle kephalometrischen Variablen in Abhängigkeit eines einzigen Merkmals (SNA) regressiert werden. Für verschiedene Ausprägungen der unabhängigen Variable werden die entsprechenden im Mittel erwarteten Werte der anderen Merkmale in einer Tabelle zusammengefasst, mit der weitere Analysen betrieben werden. Die Berücksichtigung unterschiedlicher Gesichtstypen in dieser Art führte zu dem Begriff der "floating norms". Doch leider vernachlässigt diese Methode, wie bereits erwähnt, trotzdem noch wichtige mehrdimensionale Zusammenhänge.

Im Folgenden wird ein neuer Ansatz vorgestellt, der Korrelationen zwischen allen Variablen berücksichtigt. Es werden mehrere multivariate Regressionen mit $(k-1)$ Regressoren verwendet und jede Variable auf alle anderen regressiert. Obwohl diese Methode auf den ersten Blick nicht sehr viele Gemeinsamkeiten mit der Harmoniebox hat, so werden sich gegen Ende der Arbeit einige Zusammenhänge zeigen.

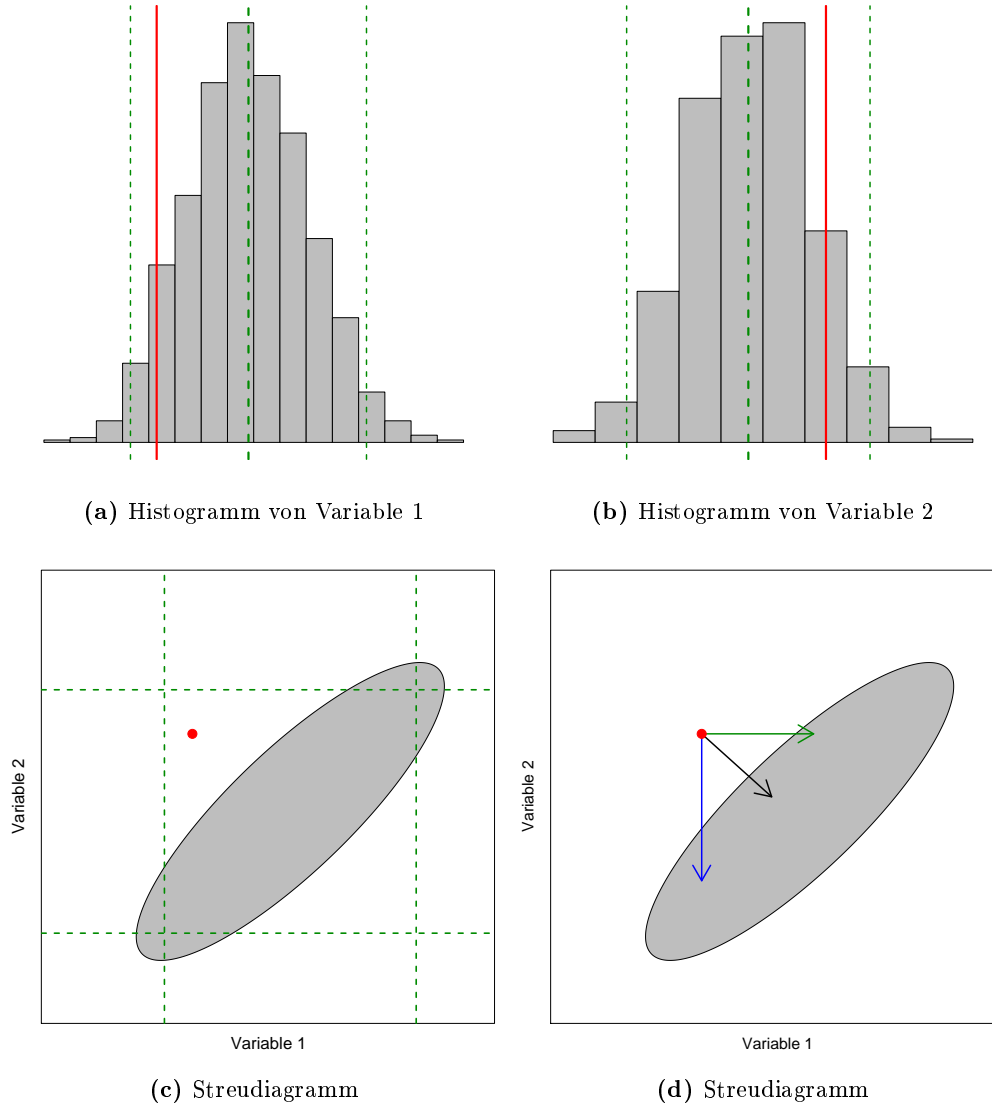


Abbildung 1.1.: Ein einführendes Beispiel zur Patientenbewertung. (a) und (b) zeigen zwei annähernd normalverteilte Variablen mittels Histogrammen. Neben den Werten eines fiktiven Patienten sind Mittelwerte und Grenzen von zwei Standardabweichungen eingezeichnet. (c) zeigt den Patienten im Streudiagramm, wobei weiterhin die 1-dimensionalen Standardabweichungsgrenzen eingezeichnet sind sowie die Punktwolke als Ellipse dargestellt wird. (d) zeigt unterschiedliche Möglichkeiten in die Ellipse zu gelangen.

Kapitel 2.

Der Multi-Harmonie-Ansatz

Wie kann man die relative Position eines Individuums mit Werten x_1, \dots, x_k zur Normgruppe mit Winkeln $\mathbf{X}_1, \dots, \mathbf{X}_k$ (aufgefasst als Datenvektoren) feststellen, die Abhängigkeitsstruktur, die dieser unterliegt, berücksichtigen und die Fragen aus Kapitel 1 beantworten?

Im ersten Abschnitt werden einige grundlegende Begriffe der linearen Regression präsentiert. Im zweiten wird der *Multi-Harmonie-Ansatz* oder kurz *MH Ansatz* in einem allgemeinen Rahmen für k Variablen vorgestellt. Die Bezeichnung rührt daher, dass es mit dem MH Ansatz möglich ist eine große Vielfalt von harmonischen Gesichtern zu beschreiben. Neben weiteren Schreibweisen und der in Schritten dargestellten Vorgehensweise beim MH Ansatz wird ein allgemeines Kriterium zur Klassifizierung eines Individuums in Bezug auf die Gruppenzugehörigkeit eingeführt.

Da viele mehrdimensionale Probleme im zweidimensionalen anschaulich dargestellt werden können, wird die hinter der Methode stehende Idee nochmals ausführlich für $k = 2$, anhand von Grafiken und eines Beispiels erläutert.

Zu guter Letzt wird der MH Ansatz auf den koreanischen Datensatz angewendet, der in Anhang A näher beschrieben ist. Die aus den Daten resultierenden Regressionsgleichungen werden genauer betrachtet, die Standardfehler der Schätzer der Parameter untersucht und einige diagnostische Grafiken gezeigt. Außerdem wird die Visualisierung der Residuen und der Winkelanpassungen und somit die Anwendung durch die Kieferorthopäden anhand von Beispielen beschrieben.

2.1. Grundlegende Begriffe von Regressionsmodellen

Seien Y_1, \dots, Y_n und F_1, \dots, F_n jeweils unabhängige, identisch verteilte \mathbb{R} -wertige Zufallsvariablen und $\mathbf{X}_1, \dots, \mathbf{X}_p \in \mathbb{R}^n$ linear unabhängige Datenvektoren. Es sei $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\mathbf{F} = (F_1, \dots, F_n)^T$ und $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$ der *Einsvektor*.

Definition 2.1. Ein *multivariates lineares Regressionsmodell* mit zufälligem Fehler \mathbf{F} und Erwartungswert $\mathbb{E}(\mathbf{F}) = \mathbf{0}$, mit $\mathbf{0} := (0, \dots, 0)^T \in \mathbb{R}^n$, ist von der Form

$$\mathbf{Y} = \beta_0 \mathbf{1} + \beta_1 \mathbf{X}_1 + \dots + \beta_p \mathbf{X}_p + \mathbf{F}. \quad (2.1)$$

\mathbf{Y} wird *abhängige Variable*, *Response*, *Zielvariable* oder *Regressand* genannt. $\mathbf{X}_1, \dots, \mathbf{X}_p$ bezeichnet man als *unabhängige Variablen*, *Prädiktoren* oder *Regressoren*. Die Koeffizienten β_0, \dots, β_p lassen sich mittels den Realisierungen y_1, \dots, y_n von Y_1, \dots, Y_n und der Methode der kleinsten Quadrate durch $\hat{\beta}_0, \dots, \hat{\beta}_p$ schätzen. Im Fall $p = 1$ bezeichnet man (2.1) als *einfaches lineares Regressionsmodell*.

Definition 2.2. Der *Fit*, der *vorhergesagte Wert* oder schlicht die *Vorhersage*, ist gegeben als

$$(\hat{Y}_1, \dots, \hat{Y}_n)^T = \hat{\mathbf{Y}} := \hat{\beta}_0 \mathbf{1} + \hat{\beta}_1 \mathbf{X}_1 + \dots + \hat{\beta}_p \mathbf{X}_p. \quad (2.2)$$

Definition 2.3. Die *Abweichung vom erwarteten bzw. vorhergesagten Wert*, kurz die *Abweichung*, oder das *Residuum*, ist definiert als

$$(r_1, \dots, r_n)^T = \mathbf{R} := \mathbf{Y} - \hat{\mathbf{Y}}. \quad (2.3)$$

Definition 2.4. Man bezeichnet

$$\hat{\sigma} := \sqrt{\frac{1}{n - (p + 1)} \sum_{i=1}^n r_i^2} \quad (2.4)$$

als den *Standardfehler der Regression*.

2.2. Der allgemeine Fall

Im Folgenden werden die Daten x_{ij} für $i = 1, \dots, n$ und $j = 1, \dots, k$ von n Normalen betrachtet und als Datenvektoren $\mathbf{X}_1, \dots, \mathbf{X}_k$ geschrieben. Für die Darstellung aller Vektoren außer einem und unter Berücksichtigung der Reihenfolge sei

$$\begin{aligned}\mathbf{X}^{(1)} &:= (\mathbf{X}_2, \dots, \mathbf{X}_k), \\ \mathbf{X}^{(j)} &:= (\mathbf{X}_1, \dots, \mathbf{X}_{j-1}, \mathbf{X}_{j+1}, \dots, \mathbf{X}_k) \quad \forall j \in \{2, \dots, k-1\}, \\ \mathbf{X}^{(k)} &:= (\mathbf{X}_1, \dots, \mathbf{X}_{k-1}).\end{aligned}\tag{2.5}$$

Des Weiteren sei

$$f_j(\mathbf{X}^{(j)}) := \beta_0^{(j)} \mathbf{1} + \beta_1^{(j)} \mathbf{X}_1 + \dots + \beta_{j-1}^{(j)} \mathbf{X}_{j-1} + \beta_j^{(j)} \mathbf{X}_{j+1} + \dots + \beta_{k-1}^{(j)} \mathbf{X}_k \tag{2.6}$$

(wobei eine analoge Definition für $\mathbf{X}^{(1)}$ und $\mathbf{X}^{(k)}$ gilt), bzw. um die Schätzung der Parameter zu verdeutlichen

$$\hat{f}_j(\mathbf{X}^{(j)}) := \hat{\beta}_0^{(j)} \mathbf{1} + \hat{\beta}_1^{(j)} \mathbf{X}_1 + \dots + \hat{\beta}_{j-1}^{(j)} \mathbf{X}_{j-1} + \hat{\beta}_j^{(j)} \mathbf{X}_{j+1} + \dots + \hat{\beta}_{k-1}^{(j)} \mathbf{X}_k. \tag{2.7}$$

Der Index (j) im Exponenten, mit $j \in \{1, \dots, k\}$, zeigt die Zugehörigkeit der Parameter β_0, \dots, β_k zu dem Regressionsmodell an, welches \mathbf{X}_j als Regressand hat. Analoge Schreibweisen gelten für \mathbf{F} und \mathbf{R} . Damit ist

$$\mathbf{X}_j = f_j(\mathbf{X}^{(j)}) + \mathbf{F}^{(j)} \tag{2.8}$$

ein multivariates Regressionsmodell der Form (2.1). Als Vorhersage setze man $(\hat{X}_{1j}, \dots, \hat{X}_{nj})^T = \hat{\mathbf{X}}_j := \hat{f}_j(\mathbf{X}^{(j)})$ und als Residuum $\mathbf{R}^{(j)} = (r_1^{(j)}, \dots, r_n^{(j)})^T$.

Wird ein zusätzliches Individuum $\mathbf{x} := (x_1, \dots, x_k)$ betrachtet, so stellt $\mathbf{x}^{(j)} := (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k)$ alle Werte außer dem j -ten dar. Die entsprechende Vorhersage des Individuums ist dann $\hat{x}_j := \hat{f}_j(\mathbf{x}^{(j)})$ und seine Abweichung gegeben durch

$$r^{(j)} := x_j - \hat{x}_j. \tag{2.9}$$

2.2.1. Vorgehensweise

Nach der Einführung in die im weiteren Verlauf verwendeten Notationen, wird sich dem Hauptteil dieser Arbeit gewidmet: der Vorstellung des MH Ansatzes.

Diese Methode stellt eine anschauliche Möglichkeit dar, Ausreißer und damit behandlungsbedürftige Individuen (Patienten) anhand der Population der Normalen zu erkennen. Des Weiteren zeigt der Ansatz inwieweit Winkelverhältnisse von denen in der Normgruppe abweichen und liefert eine Begründung für die Klassifikation.

In der Regel gilt es ein Individuum $\mathbf{x} := (x_1, \dots, x_k)$ zu betrachten und eine Diagnose bezüglich seiner Winkelkombination im Vergleich zu den Normalen zu liefern. Ausgehend von den Überlegungen und Schreibweisen aus dem vorherigen Abschnitt lässt sich der MH Ansatz in folgende Schritte zusammenfassen, welche eine leichte Implementierung am Computer ermöglichen:

1. **k Regressionsgleichungen:** Man gehe von einem linearen Zusammenhang zwischen den Winkeln aus (vgl. Anhang A.2.2), wobei jede Variable auf alle anderen regressiert wird, was zu einem System von k Regressionsgleichungen der Form (2.8) führt:

$$\begin{aligned} \mathbf{X}_1 &= f_1(\mathbf{X}^{(1)}) + \mathbf{F}^{(1)} \\ &\vdots \\ \mathbf{X}_k &= f_k(\mathbf{X}^{(k)}) + \mathbf{F}^{(k)} \end{aligned} \tag{2.10}$$

2. **Schätzung der Koeffizienten:** Die Funktionen f_1, \dots, f_k und damit die Koeffizienten $\beta_0^{(j)}, \dots, \beta_{k-1}^{(j)}$, mit $j = 1, \dots, k$, lassen sich mit Hilfe der Referenzgruppe der Normalen und deren Werten x_{i1}, \dots, x_{ik} für $i = 1, \dots, n$ mittels der Methode der kleinsten Quadrate durch $\hat{f}_1, \dots, \hat{f}_k$ bzw. $\hat{\beta}_0^{(j)}, \dots, \hat{\beta}_{k-1}^{(j)}$ schätzen. Man kann schreiben:

$$\begin{aligned} \mathbf{X}_1 &= \hat{f}_1(\mathbf{X}^{(1)}) + \mathbf{R}^{(1)} \\ &\vdots \\ \mathbf{X}_k &= \hat{f}_k(\mathbf{X}^{(k)}) + \mathbf{R}^{(k)} \end{aligned} \tag{2.11}$$

3. **Die Abweichungsgrafik:** Es wird eine Grafik erstellt, welche auf der ersten Achse den Index $(1), \dots, (k)$ (die jeweilige Response Variable) und auf der zweiten die dazu gehörenden Residuen hat. Diese Grafik wird in Zukunft als *Abweichungsgrafik*

bezeichnet. Sie entspricht im Prinzip einer Abwandlung eines *Parallelen Koordinatenplots* (siehe [ID90]). In die Abweichungsgrafik werden die Residuen der Normalen $r_1^{(j)}, \dots, r_n^{(j)}$ in die j -te Spalte eingezeichnet (für $j = 1, \dots, k$). In Abbildung 2.1a wird ein Beispiel für $k = 5$ gezeigt.

- 4. Abweichungen des Individuums:** Man bestimme nun mit Hilfe der geschätzten Funktionen $\hat{f}_1, \dots, \hat{f}_k$ die Vorhersagen $\hat{x}_1, \dots, \hat{x}_k$ des Individuums \mathbf{x} mittels

$$\begin{aligned} \hat{x}_1 &= \hat{f}_1(\mathbf{x}^{(1)}) \\ &\vdots \\ \hat{x}_k &= \hat{f}_k(\mathbf{x}^{(k)}) \end{aligned} \tag{2.12}$$

und damit die Abweichungen $r^{(1)}, \dots, r^{(k)}$ für \mathbf{x} und trage diese in der Abweichungsgrafik in der jeweiligen Spalte ein (vgl. Abbildung 2.1b).

- 5. Klassifikationsregel:** Man vergleiche spaltenweise die Abweichungen der Normalen $r_1^{(j)}, \dots, r_n^{(j)}$ mit der Abweichung $r^{(j)}$ des zusätzlichen Individuums \mathbf{x} und klassifiziere das Individuum anhand einer Regel $\lambda^{(j)}$, $j = 1, \dots, k$, (welche im nächsten Abschnitt vorgestellt wird) als normal oder behandlungsbedürftig.

- 6. Anpassung der Winkel:** Ist der Patient behandlungsbedürftig, so verändere man die Winkelwerte x_1, \dots, x_n und beginne wieder mit Schritt 3.

2.2.2. Klassifikationsregeln

Die Abweichungen stellen ein mögliches Kriterium dar, Diskrepanzen von der Normgruppe zu entdecken und zu bewerten. Betragsmäßig hohe Abweichungen deuten auf einen Punkt außerhalb der durch die Referenzgruppe definierten Punktwolke hin.

Was stellt eine große oder kleine Abweichung dar? Um ein Gefühl für die Größenordnung zu bekommen, bietet es sich an, die Abweichungen des Objektes mit der Verteilung der Abweichungen der Normalen in der Referenzgruppe zu vergleichen. Um das Vergleichen abstrakt darstellen zu können, sei an dieser Stelle ein allgemeines Kriterium angegeben, um ein Individuum als Patienten oder Normalen zu klassifizieren. Das Kriterium besagt Folgendes:

Man gebe sich jeweils einen Wert $\lambda^{(j)} \in \mathbb{R} \forall j \in \{1, \dots, k\}$ vor und überprüfe, ob

$$|r^{(j)}| \leq |\lambda^{(j)}| \forall j \in \{1, \dots, k\} \tag{2.13}$$

gilt (Erinnerung: $r^{(j)}$, $j = 1, \dots, k$, sind die Residuen/Abweichungen des zusätzlichen Individuums). Ist diese Bedingung erfüllt, so wird das Individuum als Normaler betrachtet. Ist (2.13) für mindestens ein j nicht erfüllt, so wird es als Patient klassifiziert. Diese Vorgehensweise stellt eine Vereinfachung dar, welche jedoch die Wahl von $\lambda^{(j)}$ dem Anwender und seinen individuellen Ansichten überlässt.

Beispiel 2.5. *Es seien drei offensichtliche Möglichkeiten für ein solches Kriterium kurz vorgestellt:*

- Die "max-Residuum" Regel

$$\lambda_{max}^{(j)} := \begin{cases} \max\{r_1^{(j)}, \dots, r_n^{(j)}\} & , \text{ für } r^{(j)} > 0 \\ \min\{r_1^{(j)}, \dots, r_n^{(j)}\} & , \text{ für } r^{(j)} \leq 0 \end{cases} \quad (2.14)$$

Ist die Abweichung vom vorhergesagten Wert für die j -te Variable negativ, so wird das Minimum aller Residuen der Normalen als Schranke gewählt. Ist sie positiv, so wird das Maximum gewählt.

- Die "Standardfehler" Regel

$$\lambda_{\hat{\sigma}}^{(j)}(l) := l \hat{\sigma}^{(j)}, \text{ mit } l \geq 0, \quad (2.15)$$

dabei stellt $\hat{\sigma}^{(j)}$ denjenigen Standardfehler der Regression der Gleichung aus (2.11) dar, welche X_j als Regressand hat.

- Die "Quantil" Regel

$$\lambda_q^{(j)}(l) := \begin{cases} q(\{r_1^{(j)}, \dots, r_n^{(j)}\}, 1-l) & , \text{ für } r^{(j)} > 0 \\ q(\{r_1^{(j)}, \dots, r_n^{(j)}\}, l) & , \text{ für } r^{(j)} \leq 0 \end{cases} \quad (2.16)$$

mit $l \in [0, t]$, wobei $q(M, l)$ das empirische $l \cdot 100\%$ -Quantil einer Menge $M \subset \mathbb{R}$ darstellt und t so gewählt ist, dass $\lambda_q^{(j)}(t) = \frac{1}{n} \sum_{i=1}^n r_i^{(j)}$. ■

Die Max-Residuum Regel $\lambda_{max}^{(j)}$ wird direkt durch die der Referenzgruppe zugrunde liegenden Normalen bestimmt und lässt keine zusätzlichen Modifikationen zu. Offensichtlich gilt $\lambda_{max}^{(j)} = \lambda_q^{(j)}(0)$ und $|\lambda_q^{(j)}(l)| \leq |\lambda_{max}^{(j)}| \forall l \in (0, t]$. Die Standardfehler Regel $\lambda_{\hat{\sigma}}^{(j)}(l)$ hingegen bietet die größte Flexibilität.

Es lassen sich natürlich auch weitere Regeln sowie Kombinationen aus den oben gezeigten formulieren. Wird nichts anderes erwähnt, so wird in Zukunft aus Gründen der Eindeutigkeit $\lambda_{max}^{(j)}$ verwendet. Es folgt ein kurzes Anwendungsbeispiel:

Beispiel 2.6. In Abbildung 2.1b befinden sich neben den Abweichungen des Individuums prozentuale Anzeigen, wie viele Normale solch eine oder noch extremere Abweichung aufweisen. Die Abweichung vom vorhergesagten Wert bei Variable 3 ist betragsmäßig recht hoch, jedoch gibt es in der Referenzgruppe 6 Prozent mit einer mindestens genauso großen. Es ist sofort ersichtlich, dass bei den Variablen 1,4 und 5 kein Normaler solche Abweichungen hat. Nach $\lambda_{max}^{(j)}$ (wie auch für $\lambda_q^{(j)}(l) \forall l \in [0, t]$) würde man das gezeigte Individuum als Patienten betrachten ■

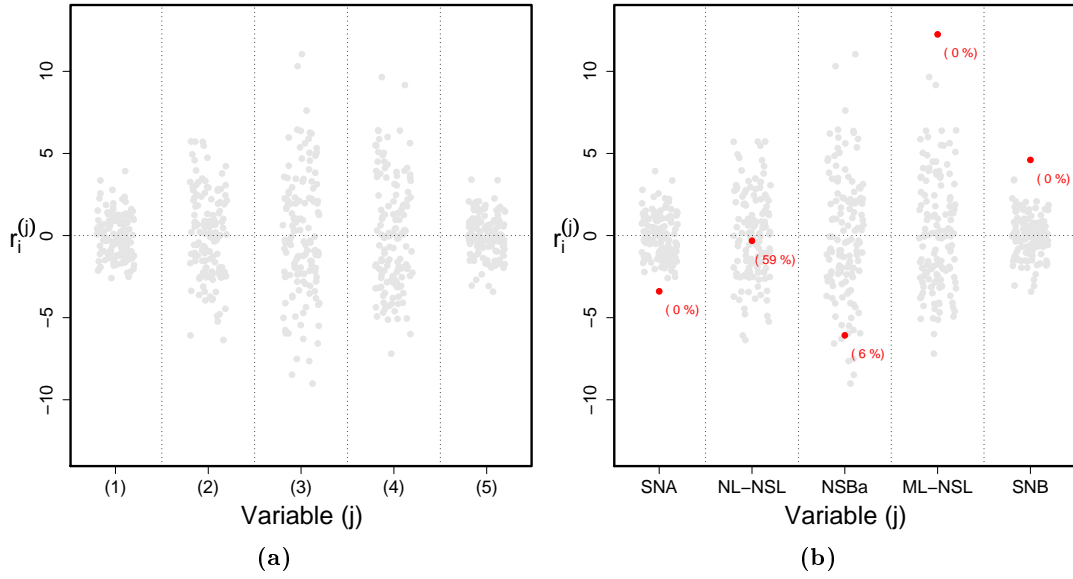


Abbildung 2.1.: Die Abweichungsgrafik. (a) Auf der ersten Achse befindet sich die (j) -te Response Variable und auf der zweiten Achse die dazugehörigen Residuen $r_i^{(j)}$. Die grauen Punkte repräsentieren die Normale. Da die erste Achse keine Information enthält, wurden die Residuen für eine bessere Darstellung mit einem unabhängigen uniform verteilten Fehler ϵ versehen. (b) Die Nummerierung des Index wurde mit Namen der Variablen ersetzt. Die zusätzlichen Punkte stellen die Abweichungen des Patienten dar. Die prozentualen Werte geben an, wie viele Normale eine solche oder noch extremere Abweichung aufweisen.

2.3. Der Ansatz für $k=2$

Die Vorgehensweise aus Abschnitt 2.2.1 wird exemplarisch für $k = 2$ vorgeführt. Angenommen es gibt nur zwei Winkel, die bei n Normalen gemessen wurden, welche sich zu den Datenvektoren \mathbf{X}_1 und \mathbf{X}_2 zusammenfassen lassen. Außerdem gibt es ein zusätzliches Individuum $\mathbf{x} = (x_1, x_2)$, von dem nicht bekannt ist, ob es einen Normalen oder Patienten repräsentiert. Entspricht die Kombination der Winkelwerte von \mathbf{x} denen der Normgruppe? Es werden die Schritte aus dem vorherigen Abschnitt angewendet:

1. k Regressionsgleichungen:

$$\begin{aligned}\mathbf{X}_1 &= \alpha_0^{(1)} \mathbf{1} + \alpha_1^{(1)} \mathbf{X}_2 + \mathbf{F}^{(1)} \\ \mathbf{X}_2 &= \alpha_0^{(2)} \mathbf{1} + \alpha_1^{(2)} \mathbf{X}_1 + \mathbf{F}^{(2)}\end{aligned}\tag{2.17}$$

Dabei repräsentiert der Index (j) im Exponenten, $j \in \{1, 2\}$, dass die Parameter α_0 , α_1 und der Fehler \mathbf{F} zu dem Modell gehören, welches die Variable \mathbf{X}_j als Response hat.

2. Schätzung der Koeffizienten: Mit Hilfe der n Normalen und der Methode der kleinsten Quadrate lassen sich die Koeffizienten $\hat{\alpha}_0^{(1)}, \hat{\alpha}_1^{(1)}$ und $\hat{\alpha}_0^{(2)}, \hat{\alpha}_1^{(2)}$ schätzen und das Ergebnis in folgenden Gleichungen zusammenfassen:

$$\begin{aligned}\mathbf{X}_1 &= \hat{\alpha}_0^{(1)} \mathbf{1} + \hat{\alpha}_1^{(1)} \mathbf{X}_2 + \mathbf{R}^{(1)} \\ \mathbf{X}_2 &= \hat{\alpha}_0^{(2)} \mathbf{1} + \hat{\alpha}_1^{(2)} \mathbf{X}_1 + \mathbf{R}^{(2)}\end{aligned}\tag{2.18}$$

In Grafik 2.2a werden die n Normalen als graue Punkte dargestellt. Geht man von den geschätzten Gleichungen (2.18) aus, so lassen sich zwei Regressionsgeraden durch die Punktwolke legen. Diese gehen immer durch den Mittelwert der Daten (Schwerpunkt der Punktwolke). Dies ergibt sich direkt als Spezialfall ($k = 2$) aus dem noch folgendem Lemma 3.1 auf Seite 25.

Des Weiteren unterscheiden sich die beiden Regressionsgeraden voneinander, solange der empirische Korrelationskoeffizient von \mathbf{X}_1 und \mathbf{X}_2 dem Betrage nach kleiner 1 ist, was im Folgenden gezeigt wird:

Definition 2.7. Der (empirische) Korrelationskoeffizient von zwei Datenvektoren $\mathbf{X} =$

$(x_1, \dots, x_n)^T$ und $\mathbf{Y} = (y_1, \dots, y_n)^T$ ist

$$\hat{\rho}_{\mathbf{XY}} := \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Lemma 2.8. Schätzt man die Koeffizienten α_0, α_1 eines Regressionsmodells $\mathbf{Y} = \alpha_0 \mathbf{1} + \alpha_1 \mathbf{X} + \mathbf{F}$ mittels der Daten x_i, y_i , für $i = 1, \dots, n$, und der Methode der kleinsten Quadrate, so sind die Koeffizienten gegeben durch $\hat{\alpha}_0 := \bar{y} - \hat{\alpha}_1 \bar{x}$ und $\hat{\alpha}_1 := \frac{s_{\mathbf{Y}}}{s_{\mathbf{X}}} \hat{\rho}_{\mathbf{XY}}$. Dabei ist $s_{\mathbf{X}} := \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$.

Beweis. Die erste Behauptung ($\hat{\alpha}_0 := \bar{y} - \hat{\alpha}_1 \bar{x}$) folgt direkt aus Lemma 3.1 auf Seite 25. Für die zweite Behauptung sei

$$0 \stackrel{!}{=} \frac{\partial}{\partial \hat{\alpha}_1} \sum_{i=1}^n (y_i - \hat{\alpha}_0 - \hat{\alpha}_1 x_i)^2 \quad (2.19)$$

$$\implies 0 = \sum_{i=1}^n y_i x_i - \hat{\alpha}_0 x_i - \hat{\alpha}_1 x_i^2 \quad (2.20)$$

$$= \sum_{i=1}^n y_i x_i - \bar{y} x_i - \hat{\alpha}_1 (x_i^2 - \bar{x} x_i) \quad (2.21)$$

$$\implies \hat{\alpha}_1 = \sum_{i=1}^n \frac{y_i x_i - \bar{y} x_i}{x_i^2 - \bar{x} x_i} \quad (2.22)$$

$$= \sum_{i=1}^n \frac{y_i - \bar{y}}{x_i - \bar{x}} \quad (2.23)$$

$$= \sum_{i=1}^n \frac{\frac{1}{n} (y_i - \bar{y})(x_i - \bar{x})}{\frac{1}{n} (x_i - \bar{x})^2} \quad (2.24)$$

$$= \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}} \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.25)$$

$$\implies \hat{\alpha}_1 = \frac{s_{\mathbf{Y}}}{s_{\mathbf{X}}} \hat{\rho}_{\mathbf{XY}} \quad (2.26)$$

□

Satz 2.9. Ist $|\hat{\rho}_{\mathbf{XY}}| < 1$, so unterscheiden sich die beiden Regressionsgeraden

$$g := \{(x, y) \in \mathbb{R}^2 \mid x = \hat{\alpha}_0^{(1)} + \hat{\alpha}_1^{(1)}y\}$$

$$h := \{(x, y) \in \mathbb{R}^2 \mid y = \hat{\alpha}_0^{(2)} + \hat{\alpha}_1^{(2)}x\}$$

Beweis. Da beide Geraden durch den Punkt (\bar{x}, \bar{y}) gehen, reicht es zu zeigen, dass die Steigungen $\hat{\alpha}_1^{(1)}$ und $1/\hat{\alpha}_1^{(2)}$ unterschiedlich sind. Aus Lemma 2.8 folgt, dass

$$\begin{aligned} \hat{\alpha}_1^{(1)} &= \frac{s_{\mathbf{X}}}{s_{\mathbf{Y}}} \hat{\rho}_{\mathbf{XY}} \quad \text{und} \quad \hat{\alpha}_1^{(2)} = \frac{s_{\mathbf{Y}}}{s_{\mathbf{X}}} \hat{\rho}_{\mathbf{XY}} \\ \implies \hat{\alpha}_1^{(1)} &= \frac{s_{\mathbf{X}}}{s_{\mathbf{Y}}} \hat{\rho}_{\mathbf{XY}} \neq \frac{s_{\mathbf{X}}}{s_{\mathbf{Y}}} \frac{1}{\hat{\rho}_{\mathbf{XY}}} = 1/\hat{\alpha}_1^{(2)} \end{aligned}$$

□

3. Die Abweichungsgrafik: Die Residuen der Regression werden in eine Abweichungsgrafik mit zwei Spalten eingetragen. In die erste kommen $r_1^{(1)}, \dots, r_n^{(1)}$ und in die zweite $r_1^{(2)}, \dots, r_n^{(2)}$. Es sei nochmals angemerkt, dass die horizontale Achse keine Information beinhaltet. Die Residuen der Normalen werden der Übersicht halber wieder mit einem unabhängigen uniform verteilten Fehler ϵ versehen (siehe Abb. 2.2b).

4. Abweichungen des Individuums: Mit den geschätzten Koeffizienten lassen sich die Vorhersagen für $\mathbf{x} = (x_1, x_2)$ bestimmen:

$$\begin{aligned} \hat{x}_1 &= \hat{\alpha}_0^{(1)} + \hat{\alpha}_1^{(1)}x_2 \\ \hat{x}_2 &= \hat{\alpha}_0^{(2)} + \hat{\alpha}_1^{(2)}x_1 \end{aligned} \tag{2.27}$$

Die Abweichungen für \mathbf{x} resultieren als $r^{(1)} := x_1 - \hat{x}_1$ und $r^{(2)} := x_2 - \hat{x}_2$ (man vgl. Abb. 2.2a). Die Größe dieser Differenzen wird in die bestehende Abweichungsgrafik eingetragen.

5. Klassifikationsregel: Aus der Abweichungsgrafik 2.2b ist ersichtlich, dass die Abweichungen von \mathbf{x} betragsmäßig deutlich größer sind als die der Normalen. Da die Schranke $\lambda_{max}^{(j)}$ deutlich überschritten wird, würde man \mathbf{x} somit als Patienten betrachten.

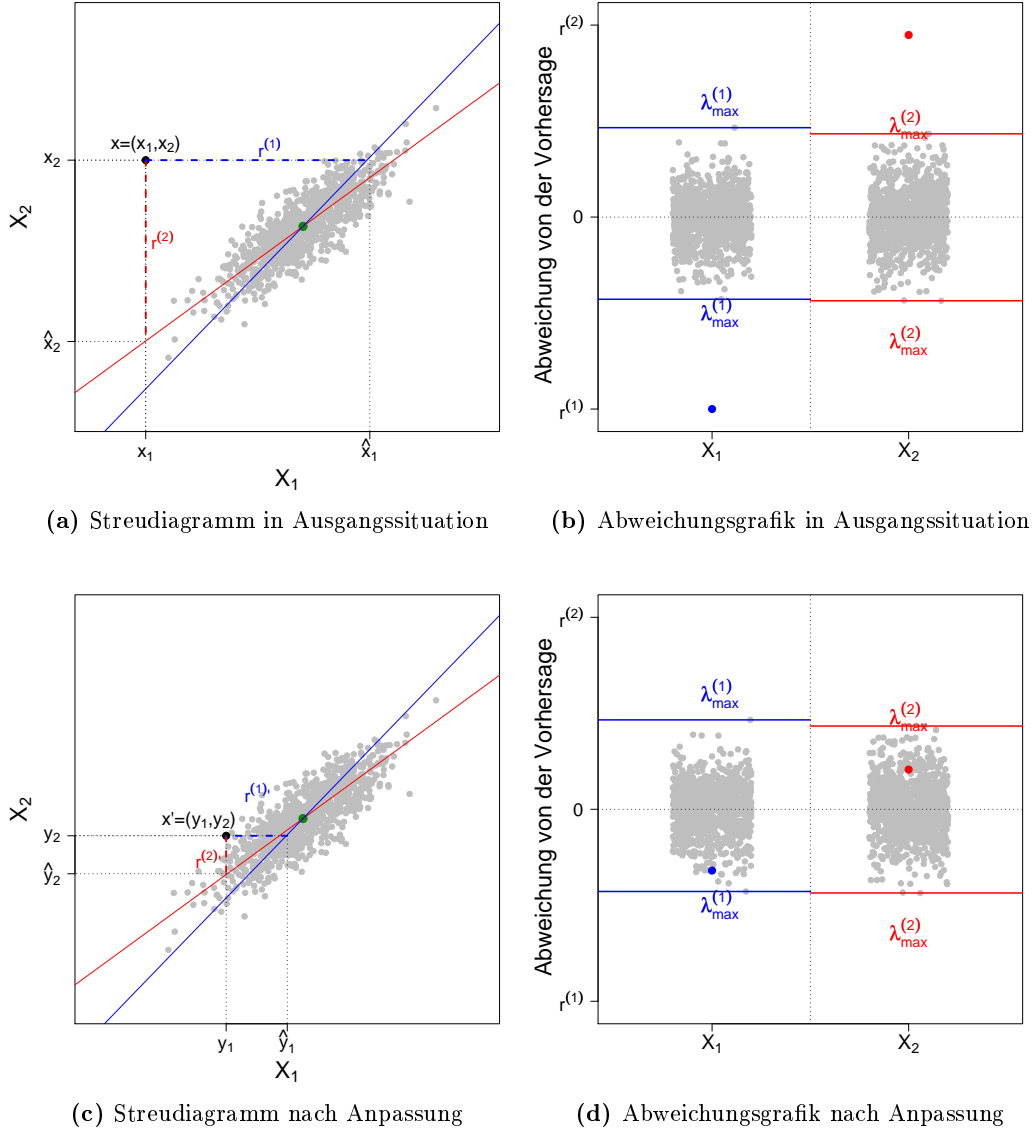


Abbildung 2.2.: Das Modell für $p = 2$. (a) zeigt die Position eines Individuums im Vergleich zu den Normalen. Es sind die Regressionsgeraden und die entsprechenden Residuen eingezeichnet. (b) zeigt die dazugehörige Abweichungsgrafik mit Grenzen für $\lambda_{max}^{(j)}$. (c) und (d) stellen die neue Situation nach Änderung der Winkelwerte dar.

6. Anpassung der Winkel: Die Kombination der Winkelwerte x_1, x_2 passt somit nicht zu den Winkelkombinationen der Norm. Man ändere die Werte \mathbf{x} und führe die Schritte 3 bis 5 mit den neuen Werten y_1, y_2 aus. Das Vorgehen ist nochmals in Abb. 2.2c dargestellt. Aus der neuen Abweichungsgrafik 2.2d ist ersichtlich, dass die Abweichungen nun innerhalb der durch $\lambda_{max}^{(j)}$ definierten Schranken sind. Das Individuum wird als normal klassifiziert.

2.4. Anwendung auf den Datensatz

Basierend auf der Vorgehensweise aus Abschnitt 2.2.1 wird der MH Ansatz auf den koreanischen Datensatz angewendet. Dieser besteht aus zwei Gruppen (134 Normale und 86 Patienten) sowie den kephalometrischen Winkeln SNA, NL-NSL, NSBa, ML-NSL und SNB. Eine kurze Beschreibung der Variablen und einige deskriptive Statistiken findet man in Anhang A.

2.4.1. Regressionen

Das Augenmerk wird zuerst auf die zugrunde liegenden Regressionen gelegt, wobei die Vorgehensweise für den Fall $k = 5$ Schritt für Schritt abgearbeitet wird. Man setze $SNA =: X_1$, $NL-NSL =: X_2$, $NSBa =: X_3$, $ML-NSL =: X_4$ und $SNB =: X_5$.

1. k Regressionsgleichungen:

Beim Aufstellen eines Systems der Form (2.10)(siehe Seite 10) betrachtet man somit folgende Gleichungen:

$$\begin{array}{rclclcl}
 SNA & = & \beta_0^{(1)} + \beta_1^{(1)}NL-NSL & + \beta_2^{(1)}NSBa & + \beta_3^{(1)}ML-NSL & + \beta_4^{(1)}SNB & + F^{(1)} \\
 NL-NSL & = & \beta_0^{(2)} + \beta_1^{(2)}SNA & + \beta_2^{(2)}NSBa & + \beta_3^{(2)}ML-NSL & + \beta_4^{(2)}SNB & + F^{(2)} \\
 NSBa & = & \beta_0^{(3)} + \beta_1^{(3)}SNA & + \beta_2^{(3)}NL-NSL & + \beta_3^{(3)}ML-NSL & + \beta_4^{(3)}SNB & + F^{(3)} \\
 ML-NSL & = & \beta_0^{(4)} + \beta_1^{(4)}SNA & + \beta_2^{(4)}NL-NSL & + \beta_3^{(4)}NSBa & + \beta_4^{(4)}SNB & + F^{(4)} \\
 SNB & = & \beta_0^{(5)} + \beta_1^{(5)}SNA & + \beta_2^{(5)}NL-NSL & + \beta_3^{(5)}NSBa & + \beta_4^{(5)}ML-NSL & + F^{(5)}
 \end{array} \quad (2.28)$$

2. Schätzung der Koeffizienten: Man verwendet die Normalen, um die Regressionskoeffizienten zu schätzen und so die als harmonisch klassifizierten Beziehungen zwischen den Winkeln zu beschreiben:

$$\begin{array}{rcllclcl}
\widehat{\text{SNA}} & = & 23.15 & +0.10 \text{ NL-NSL} & -0.06 \text{ NSBa} & +0.06 \text{ ML-NSL} & +0.81 \text{ SNB} \\
\widehat{\text{NL-NSL}} & = & 16.13 & +0.38 \text{ SNA} & +0.12 \text{ NSBa} & +0.20 \text{ ML-NSL} & -0.74 \text{ SNB} \\
\widehat{\text{NSBa}} & = & 184.90 & -0.54 \text{ SNA} & +0.25 \text{ NL-NSL} & +0.00 \text{ ML-NSL} & -0.13 \text{ SNB} \\
\widehat{\text{ML-NSL}} & = & 79.48 & +0.41 \text{ SNA} & +0.33 \text{ NL-NSL} & +0.00 \text{ NSBa} & -1.07 \text{ SNB} \\
\widehat{\text{SNB}} & = & 27.29 & +0.74 \text{ SNA} & -0.17 \text{ NL-NSL} & -0.01 \text{ NSBa} & -0.15 \text{ ML-NSL}
\end{array} \quad (2.29)$$

Tabelle 2.1 zeigt die Standardfehler der Koeffizienten, sowie die Signifikanzniveaus der Hypothese $H_0 : \hat{\beta}_i^{(j)} = 0$ für $j = 1, \dots, 5$ und $i = 0, \dots, 4$.

Response	Intercept	SNA	NL-NSL	NSBa	ML-NSL	SNB
SNA	7.427 **	—	0.043 *	0.030 *	0.034	0.058 ***
NL-NSL	15.249	0.172 *	—	0.060 *	0.065 **	0.170 ***
NSBa	15.020 ***	0.249 *	0.125 *	—	0.098	0.264
ML-NSL	18.708 ***	0.225	0.111 **	0.080	—	0.219 ***
SNB	6.983 ***	0.053 ***	0.040 ***	0.029	0.030 ***	—
Signifikanzniveau: <0.001 ***, 0.01 **, 0.05 *						

Tabelle 2.1.: Standardfehler und Signifikanzniveaus der multivariaten Regression. Die Spalte Response zeigt die abhängige Variable (der Regressand) einer jeden Regressionsgleichung. Die restlichen Spalten stehen für die Standardfehler der jeweiligen Koeffizienten. Jede Zeile repräsentiert somit eine Regressionsgleichung. Ein Großteil der Parameter ist auf einem Niveau von $p < 0.05$ signifikant von Null verschieden.

Definition 2.10. Betrachtet man ein Regressionsmodell der Form $\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X}_1 + \dots + \beta_p \mathbf{X}_p + \mathbf{F}$, dann ist das **Bestimmtheitsmaß** (oder **multipler Korrelationskoeffizient**) definiert als

$$\mathcal{R}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.30)$$

\mathcal{R}^2 ist ein Maß für die Güte der Anpassung und es gilt immer $0 \leq \mathcal{R}^2 \leq 1$. Ein Kritikpunkt des Bestimmtheitsmaßes ist, dass es mit steigender Anzahl von Regressoren per Definition niemals sinken kann, auch wenn die zusätzlichen Variablen für die Regression völlig ungeeignet sind. Man betrachtet in solchen Fällen auch das korrigierte Bestimmtheitsmaß.

Definition 2.11. Als *korrigiertes Bestimmtheitsmaß* bezeichnet man

$$\bar{\mathcal{R}}_c^2 = 1 - (1 - \mathcal{R}^2) \frac{n-1}{n-p-1}. \quad (2.31)$$

Die beiden Bestimmtheitsmaße und Standardfehler der multivariaten Regression sind in Tabelle 2.2 angegeben. Das \mathcal{R}^2 ist bei allen Gleichungen, bis auf die mit NSBa als Response, relativ hoch und geht in diesen Fällen mit niedrigen Standardfehlern der Regression einher. Weil das korrigierte Bestimmtheitsmaß in der Tabelle in etwa \mathcal{R}^2 entspricht, gibt es keinen Grund zu glauben, dass das hohe \mathcal{R}^2 ausschließlich durch die Anzahl an Regressoren zustande kommt.

(j)	Response	\mathcal{R}^2	$\bar{\mathcal{R}}_c^2$	$\hat{\sigma}^{(j)}$
1	SNA	0.74	0.73	1.31
2	NL-NSL	0.47	0.46	2.61
3	NSBa	0.27	0.25	3.78
4	ML-NSL	0.48	0.46	3.40
5	SNB	0.82	0.82	1.26

Tabelle 2.2.: Bestimmtheitsmaße und Standardfehler der multivariaten Regression. Die Spalte Response steht für die Regressionsgleichung, welche die entsprechende Variable als Regressand hat. Das korrigierte Bestimmtheitsmaß $\bar{\mathcal{R}}_c^2$ entspricht in etwa \mathcal{R}^2 .

Es folgt eine visuelle Überprüfung der Regressionen und Residuen in Abbildung 2.3. Keine der Grafiken zeigt größere Auffälligkeiten.

2.4.2. Visualisierung der Residuen

Im Folgenden werden die aus den geschätzten Regressionsgleichungen gewonnenen Residuen sowie deren Darstellung in der Abweichungsgrafik betrachtet.

3. Die Abweichungsgrafik: Mit Hilfe der Gleichungen (2.29) lassen sich die vorhergesagten Werte und nach (2.3) (Seite 8) die Abweichungen bestimmen, welche in eine neue Abweichungsgrafik eingetragen werden.

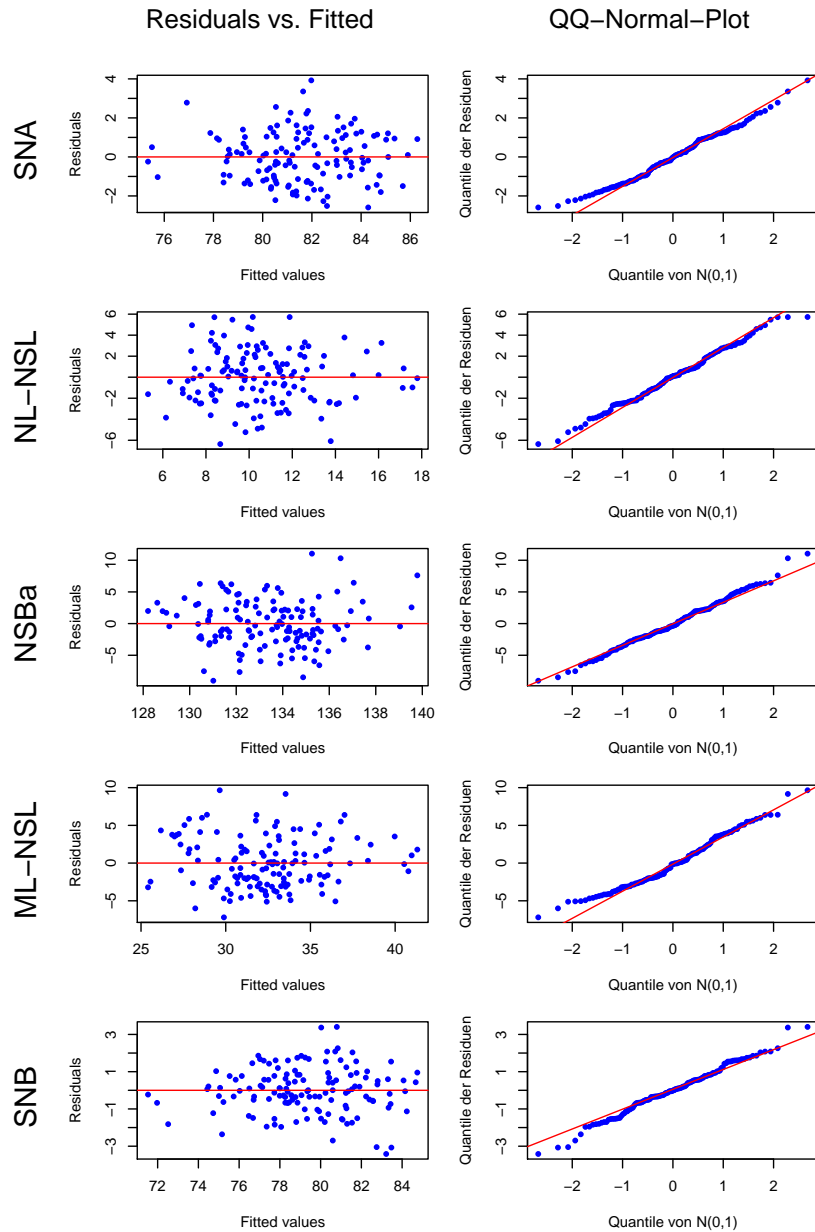


Abbildung 2.3.: Diagnostische Grafiken für die multivariate Regression. Die erste Spalte enthält die Residuen aufgetragen gegen die gefitteten (vorhergesagten) Werte, um die Annahme der konstanten Varianz der Residuen zu überprüfen. In der zweiten Spalte befinden sich Grafiken, die die Verteilung der Residuen mit einem QQ-Plot auf Normalverteilung untersuchen.

$$\begin{aligned}
 R_1 &= \text{SNA} - \widehat{\text{SNA}} \\
 R_2 &= \text{NL-NSL} - \widehat{\text{NL-NSL}} \\
 R_3 &= \text{NSBa} - \widehat{\text{NSBa}} \\
 R_4 &= \text{ML-NSL} - \widehat{\text{ML-NSL}} \\
 R_5 &= \text{SNB} - \widehat{\text{SNB}}
 \end{aligned} \tag{2.32}$$

Grafik 2.1a auf Seite 13 ist diese Abweichungsgrafik aus dem Datensatz.

4. Abweichungen des Individuums: Zur Visualisierung der Methode denke man sich als Beispiel ein fiktives Individuum, dessen Winkelwerte, Vorhersagen und Abweichungen in Tabelle 2.3 gegeben sind. In Abbildung 2.4a wird die Abweichungsgrafik für dieses Individuum gezeigt, wobei die Zahlen in Höhe der Null die vorhergesagten Werte darstellen.

Winkel	Wert	Vorhersage	Abweichung
SNA	84	87.3	-3.3
NL-NSL	6	6.3	-0.3
NSBa	124	130	-6
ML-NSL	37	24.7	12.3
SNB	86	81.3	5.7

***Tabelle 2.3.: Winkelwerte eines fiktiven Individuums.** Neben den Winkelwerten des fiktiven Patienten sind seine Vorhersagen und daraus resultierenden Abweichungen eingetragen, welche sich mittels der Gleichungen (2.29) auf Seite 19 nachrechnen lassen.*

5. Klassifikationsregel: Es ist ersichtlich, dass einige Winkel zu große Abweichungen haben und das Individuum somit unter der Verwendung von $\lambda_{max}^{(j)}$ als behandlungsbedürftig klassifiziert wird.

2.4.3. Visualisierung von Anpassungen

6. Anpassung der Winkel: Wie sollte ein Kieferorthopäde vorgehen, um diese offensichtlichen Diskrepanzen auszugleichen? An dieser Stelle werden keine Handlungsempfehlungen, sondern unterschiedliche Möglichkeiten dargestellt, wie man den Patienten (mit

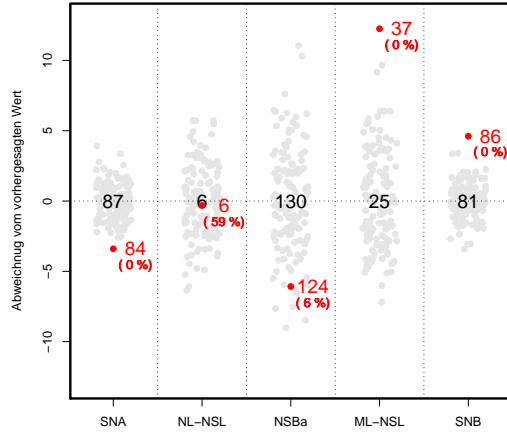
Werten x_1, \dots, x_5) in Richtung Referenzgruppe und harmonischer Winkelkombination bewegt.

Ist $r^{(j)} > 0$ und $|r^{(j)}| > |\lambda_{max}^{(j)}|$, so ist die Abweichung des Patienten im Vergleich zu den Abweichungen der Normalen als zu groß anzusehen. Da $r^{(j)} = x_j - \hat{x}_j$ ist, lässt es sich auch so formulieren, dass der Winkel x_j gegeben allen anderen Variablen zu groß ist. Die intuitive Vorgehensweise wäre es, diesen Winkel mit kieferorthopädischen Maßnahmen zu verkleinern und die neuen Winkelwerte in einer Abweichungsgrafik zu betrachten. Eine analoge Aussage lässt sich natürlich auch für $r^{(j)} < 0$ mit $|r^{(j)}| > |\lambda_{max}^{(j)}|$ treffen.

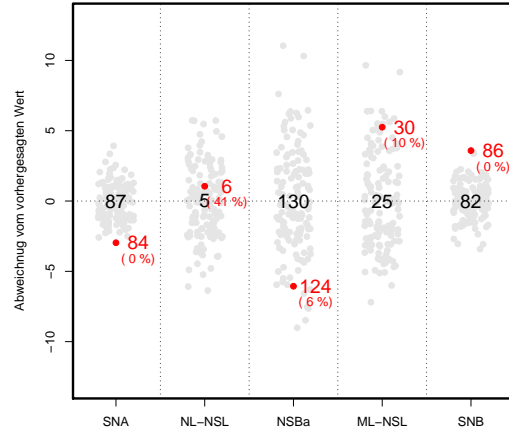
Beispiel 2.12. *Der Patient aus Tabelle 2.3 wird mit der dazugehörigen Abweichungsgrafik in Abbildung 2.4a betrachtet. Man könnte folgende Anpassungen vornehmen:*

- *Es wird eine Verkleinerung von ML-NSL um 7 Grad gewählt und damit ein Zustand von 30 Grad angestrebt. Da jede Variable in jedem der fünf Regressionsmodelle (2.28) (Seite 18) auftaucht, ändern sich auch die Abweichungen für die anderen Variablen, da sie nun neue Vorhersagen haben (man achte auf die schwarzen Werte bei ML-NSL und SNB). Die neue Situation wird in Abbildung 2.4b dargestellt. Das Individuum wäre weiterhin behandlungsbedürftig.*
- *Man gehe wieder vom ursprünglichen Patienten aus und versuche eine Reduzierung von SNB um 3 Grad. Das daraus resultierende Ergebnis findet man in Abbildung 2.4c. Als Konsequenz befinden sich nun neben SNB auch alle anderen Winkel deutlich innerhalb der durch die Referenzgruppe dargestellten grauen Punktwolke und man würde das Individuum nach $\lambda_{max}^{(j)}$ als normal betrachten.*
- *Als ein letztes Beispiel sei auf Abbildung 2.4d verwiesen, in der ML-NSL um 4 und SNB um 2 Grad verringert wurden. Der Leser überzeuge sich davon, dass mit $\lambda_{max}^{(j)}$ wie auch mit $\lambda_q^{(j)}(0.05)$ das Individuum als normal klassifiziert wird ■*

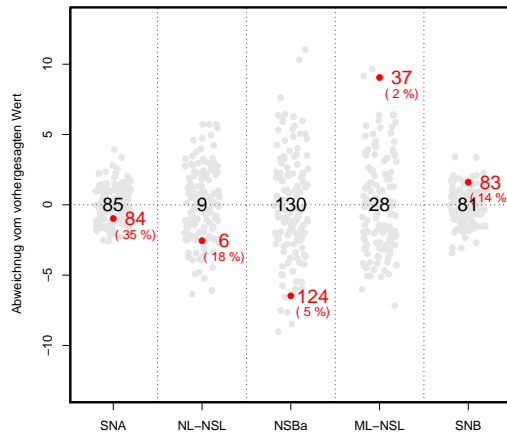
Verschiedene Anpassungen können also die Werte des Patienten in ein harmonisches Verhältnis bringen. Dabei sind die Kieferorthopäden nicht darauf beschränkt dies mittels einzelner Winkel zu erreichen, sondern können auf jede beliebige Kombination zurückgreifen.



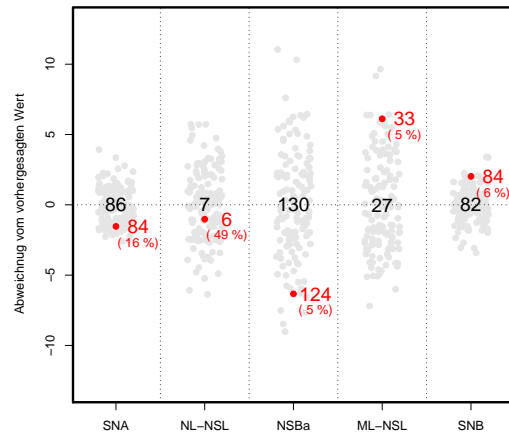
(a) Abweichungsgrafik des Individuums



(b) Anpassung von ML-NSL



(c) Anpassung von SNB



(d) Anpassung von ML-NSL und SNB

Abbildung 2.4.: Das Modell in Anwendung. (a) zeigt die Abweichungsgrafik eines Beispielpatienten. Neben seinen Winkelwerten werden die vorhergesagten Werte auf Höhe der Null angezeigt. Die restlichen Grafiken (b), (c) und (d) zeigen neue Situationen des Patienten nach verschiedenen Anpassungen.

Kapitel 3.

Interpretation und Eigenschaften

In diesem Kapitel gilt es eine anschauliche geometrische Interpretation der allgemeinen MH Methode im Raum der Variablen zu geben. Während dieser Interpretation werden einige Eigenschaften aufgedeckt und formal bewiesen. Des Weiteren sollen die möglichen beobachtbaren Fälle der Patientenklassifikation in der Abweichungsgrafik am Beispiel $k = 2$ aufgezeigt und diskutiert werden. Im Anschluss wird das Verhalten des MH Ansatzes beim Ändern der Winkelwerte und der Zusammenhang mit den Koeffizienten der Regressionsgleichungen kurz erläutert.

3.1. Geometrische Betrachtung

Was genau passiert im k -dimensionalen Variablenraum bei Anwendung der MH Methode?

Eine erste geometrische Anschauung wurde in Abschnitt 2.3 für $k = 2$ gezeigt. Auf Seite 17 zeigt Abbildung 2.2b, dass entsprechend des Gleichungssystems (2.17) (S. 14) zwei Regressionsgeraden durch die Punktwolke der Normalen gelegt werden. Diese Geraden gehen immer durch den Schwerpunkt der Daten.

Im Raum mit k Variablen hat man aufgrund von System (2.10) (siehe Seite 10) insgesamt k Regressionsgleichungen, bei denen jeweils eine Variable in Abhängigkeit von $k - 1$ Variablen beschrieben wird. Entsprechend handelt es sich nicht um Geraden, sondern um $(k-1)$ -dimensionale Hyperebenen. Alle diese Hyperebenen gehen durch den Schwerpunkt, was im Folgendem gezeigt wird:

Lemma 3.1. *Jede Regressionshyperebene $\{(x_1, \dots, x_k) \in \mathbb{R}^k \mid x_k = \hat{\alpha}_0 \mathbf{1} + \hat{\alpha}_1 x_1 + \dots + \hat{\alpha}_{k-1} x_{k-1}\}$, welche aus einem linearen Regressionsmodell der Form $\mathbf{X}_k = \alpha_0 \mathbf{1} +$*

$\alpha_1 \mathbf{X}_1 + \dots + \alpha_{k-1} \mathbf{X}_{k-1} + \mathbf{F}$ resultiert, dessen Parameter mit Hilfe der Daten x_{i1}, \dots, x_{ik} , $i = 1, \dots, n$, und der Methode der Kleinsten Quadrate geschätzt wurden, geht durch den Schwerpunkt $(\bar{x}_1, \dots, \bar{x}_k)$.

Beweis. Bei Verwendung der Methode der Kleinsten Quadrate gilt:

$$\sum_{i=1}^n (x_{ik} - \hat{\alpha}_0 - \hat{\alpha}_1 x_{i1} - \dots - \hat{\alpha}_{k-1} x_{i(k-1)})^2 = \min_{\alpha_0, \dots, \alpha_{k-1}} \sum_{i=1}^n (x_{ik} - \alpha_0 - \alpha_1 x_{i1} - \dots - \alpha_{k-1} x_{i(k-1)})^2$$

Daraus folgen die Normalgleichungen $\frac{\partial D}{\partial \hat{\alpha}_j} \stackrel{!}{=} 0$, für $j = 0, \dots, k$ und insbesondere

$$\frac{\partial D}{\partial \hat{\alpha}_0} \stackrel{!}{=} 0 \quad (3.1)$$

$$\implies 0 = -2 \sum_{i=1}^n (x_{ik} - \hat{\alpha}_0 - \hat{\alpha}_1 x_{i1} - \dots - \hat{\alpha}_{k-1} x_{i(k-1)}) \quad (3.2)$$

$$\implies \sum_{i=1}^n x_{ik} = \sum_{i=1}^n \hat{\alpha}_0 + \sum_{i=1}^n \hat{\alpha}_1 x_{i1} + \dots + \sum_{i=1}^n \hat{\alpha}_{k-1} x_{i(k-1)} \quad (3.3)$$

$$\implies \frac{1}{n} \sum_{i=1}^n x_{ik} = \hat{\alpha}_0 + \hat{\alpha}_1 \frac{1}{n} \sum_{i=1}^n x_{i1} + \dots + \hat{\alpha}_{k-1} \frac{1}{n} \sum_{i=1}^n x_{i(k-1)} \quad (3.4)$$

$$\implies \bar{x}_k = \hat{\alpha}_0 + \hat{\alpha}_1 \bar{x}_1 + \dots + \hat{\alpha}_{k-1} \bar{x}_{k-1} \quad (3.5)$$

□

Es gilt folgendes nützliches Korollar:

Korollar 3.2. Jede Regressionshyperebene $\{(x_1, \dots, x_k) \in \mathbb{R}^k \mid x_k = \hat{\alpha}_0 \mathbf{1} + \hat{\alpha}_1 x_1 + \dots + \hat{\alpha}_{k-1} x_{k-1}\}$ der zentrierten Daten $\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_k$ (mit $\tilde{\mathbf{X}}_j := \mathbf{X}_j - \bar{\mathbf{x}}_j \forall j \in \{1, \dots, k\}$, wobei $\bar{\mathbf{x}}_j := (\bar{x}_j, \dots, \bar{x}_j)^T \in \mathbb{R}^n$), welche aus einem linearen Regressionsmodell der Form $\tilde{\mathbf{X}}_k = \hat{\alpha}_0 \mathbf{1} + \hat{\alpha}_1 \tilde{\mathbf{X}}_1 + \dots + \hat{\alpha}_{k-1} \tilde{\mathbf{X}}_{k-1} + \mathbf{F}$ resultiert, dessen Parameter mit Hilfe der Daten x_{i1}, \dots, x_{ik} , $i = 1, \dots, n$, und der Methode der Kleinsten Quadrate geschätzt wurden, geht durch den Nullpunkt.

Beweis. Weil der Nullpunkt der Schwerpunkt der zentrierten Daten ist, folgt die Behauptung direkt aus Lemma 3.1. □

Unter gewissen Voraussetzungen, die beim System (2.10) auf Seite 10 fast immer erfüllt sind, schneiden sich diese Hyperebenen ausschließlich im Schwerpunkt $(\bar{x}_1, \dots, \bar{x}_k)$. Es wird dafür ein kurzes Kriterium angegeben:

Satz 3.3. *Hat man ein System von linearen Regressionsgleichungen wie in (2.10), so schneiden sich die durch die Gleichungen definierten Hyperebenen*

$$\begin{aligned} \{(x_1, \dots, x_k) \in \mathbb{R}^k \mid x_1 &= \hat{\beta}_0^{(1)} + \hat{\beta}_1^{(1)}x_2 + \dots + \hat{\beta}_{k-1}^{(1)}x_k\} \\ &\vdots \\ \{(x_1, \dots, x_k) \in \mathbb{R}^k \mid x_k &= \hat{\beta}_0^{(k)} + \hat{\beta}_1^{(k)}x_1 + \dots + \hat{\beta}_{k-1}^{(k)}x_{k-1}\} \end{aligned} \quad (3.6)$$

genau dann in einem Punkt, wenn die Matrix $\mathbf{A} = (a_{ji})$, mit $i = 1, \dots, k$, $j = 1, \dots, k$ und Einträgen

$$a_{ji} = \begin{cases} 1 & i = j \\ -\hat{\beta}_{i-1}^{(j)} & i > j \\ -\hat{\beta}_i^{(j)} & i < j \end{cases}$$

vollen Rang hat. Dieser Punkt ist der Schwerpunkt $(\bar{x}_1, \dots, \bar{x}_k)$ der Daten.

Beweis. Das System von Regressionshyperebenen lässt sich nach Umsortieren der einzelnen Summanden in Koordinatendarstellung schreiben als

$$\begin{aligned} \beta_0^{(1)} &= x_1 - \beta_1^{(1)}x_2 - \dots - \beta_{k-2}^{(1)}x_{k-1} - \beta_{k-1}^{(1)}x_k \\ &\vdots \\ \beta_0^{(k)} &= -\beta_1^{(k)}x_1 - \beta_2^{(k)}x_2 - \dots - \beta_{k-1}^{(k)}x_{k-1} + x_k \end{aligned} \quad (3.7)$$

Dieses inhomogene Gleichungssystem lässt sich mit der Matrix \mathbf{A} zusammenfassen zu $\mathbf{A}(x_1, \dots, x_k)^T = (\beta_0^{(1)}, \dots, \beta_0^{(k)})^T$. Offensichtlich gilt, dass dieses genau dann nur eine Lösung hat, wenn \mathbf{A} vollen Rang besitzt. Somit schneiden sich die Hyperebenen dann in einem Punkt. Da nach Lemma 3.1 alle Hyperebenen zumindest den Schwerpunkt enthalten, muss dies der Schnittpunkt sein. \square

Somit schneiden sich diese Hyperebenen unter diesen fast immer erfüllten und leicht zu überprüfenden Voraussetzungen ausschließlich im Schwerpunkt $(\bar{x}_1, \dots, \bar{x}_k)$. Dies führt unmittelbar zu folgendem Korollar:

Korollar 3.4. *Schneiden sich die Hyperebenen aus (3.6) ausschließlich in einem Punkt, so ergibt sich beim Weglassen einer Hyperebene als neue Schnittmenge eine Gerade.*

Beweis. Angenommen die Hyperebenen schneiden sich in einem Punkt. Dann hat nach Satz 3.3 die Matrix \mathbf{A} vollen Rang (Rang k). Wird eine Hyperebene weggelassen (und somit eine Zeile der Matrix), so hat die Matrix nur noch Rang $k - 1$. Die Dimension des inhomogenen Lösungsraumes ergibt sich somit als $k - (k - 1) = 1$, was einer Gerade entspricht. \square

Zurück zur geometrischen Interpretation: Die in der Abweichungsgrafik gezeigten Punkte stellen die Differenzen der Punkte im Variablenraum von der jeweiligen Hyperebene in Richtung des dazugehörigen Regressanden dar (man vgl. Abb. 2.2a). Der Schwerpunkt ist somit der einzige Punkt, für den gilt $r^{(1)} = \dots = r^{(k)} = 0$ und der die Regel $\lambda_{\hat{\sigma}}^{(j)}(0)$ erfüllt (vgl. (2.15) auf Seite 12).

Eine Klassifikationsregel $\lambda^{(j)}$ zur Klassifikation einer Person in "normal" oder "behandlungsbedürftig" nach (2.13) (S. 11) entspricht allgemein einem Streifen, dessen Breite von der Wahl von $\lambda^{(j)}$ abhängt, der über die j -te Hyperebene gelegt wird. Für $\lambda_{max}^{(j)}$ ist dieses Band gerade so breit, dass es die Punktwolke der Normalen komplett umschließt. Insgesamt gibt es k dieser Streifen. Man betrachte dafür Abbildung 3.1a für $k = 2$.

Der Schnitt dieser k Streifen entspricht im Mehrdimensionalen einem Parallelotop, was im Falle von $\lambda_{max}^{(j)}$ die Punktwolke der Normalen umhüllt. In der Grafik ist an der Stelle, an der sich die beiden farbigen Bänder überdecken, ein gräuliches Parallelogramm zu erkennen. Alle Punkte innerhalb dieser Schnittmenge werden von der Regel $\lambda_{max}^{(j)}$ als normal klassifiziert.

Offensichtlich gibt es nach Abbildung 3.1a Punkte, die außerhalb der Normwolke liegen und als normal klassifiziert werden (zum Beispiel Punkt D). Die Wahl eines kleineren $\lambda^{(j)}$ verringert die Chance, auf solche Individuen zu treffen - jedoch hat dies auch zur Konsequenz, dass nicht mehr alle Punkte der Normwolke überdeckt und einige Normale als Patienten betrachtet werden. Die Sensitivität und Spezifität in Bezug auf die Patientwertung hängt somit von $\lambda^{(j)}$ ab.

3.2. Möglichkeiten der Klassifikation

Die Klassifikation eines Punktes hängt mit dessen Lage und der Größe des Parallelotops zusammen (welches von $\lambda^{(j)}$ abhängt). Es wird ein Beispiel für $k = 2$ betrachtet und man identifiziere zur Illustration die Punktwolke der Normalen mit einem Normalellipsoid.

In Abbildung 3.1 wird der Zusammenhang zwischen Abweichungsgrafik (Abb. 3.1b) und Variablenraum (Abb. 3.1a) hergestellt. In beiden Grafiken sind farbige Bänder, $\lambda_{max}^{(j)}$ entsprechend, eingezeichnet. In Bezug auf die Klassifikation gibt es vier Fälle, die in der Tabelle 3.1 zusammengefasst werden. Es folgen einige Beispiele, die in der Tabelle und zum Teil in den genannten Grafiken wiederzufinden sind.

		klassifiziert als	
		normal	Patient
Person ist	normal	korrekt (B)	Fehlklassifikation (E)
	Patient	Fehlklassifikation (D)	korrekt (A,C)

Tabelle 3.1.: Möglichkeiten der Klassifizierung. Eine Vier-Felder-Tafel, die in der Zeile den tatsächlichen Zustand eines Individuums anzeigt und in der jeweiligen Spalte, wie es durch den MH Ansatz klassifiziert wurde.

Beispiel 3.5. Es wird sich auf die oben genannten Grafiken bezogen:

A Punkt A liegt außerhalb der durch die Ellipse dargestellten Referenzgruppe und dementsprechend weit von den farbigen Streifen entfernt. Diese Tatsache spiegelt sich in der Abweichungsgrafik durch eine sehr hohe (im Vergleich zur Referenzgruppe) negative Abweichung bei Variable 1 und ein sehr hohes positives Residuum bei Variable 2 wider.

B Dieser Punkt liegt deutlich innerhalb der Ellipse, relativ nah am Zentrum (dem Schnittpunkt der Geraden) und besitzt sehr kleine Abweichungen.

C Punkt C liegt außerhalb der Punktwolke. Er befindet sich zwar inmitten des roten Bandes und Variable 1 hat deswegen eine relativ kleine Abweichung (zumindest gibt

es bei den Normalen noch extremere), jedoch liegt er außerhalb des blauen Bandes. Da er sich außerhalb des Schnittes der beiden Bänder befindet, würde man das Individuum als Patient klassifizieren.

D Dieser Punkt wird, obwohl er außerhalb der Wolke liegt, als in der Wolke liegend klassifiziert.

E Der letzte Fall, dass ein Individuum normal ist, aber als Patient klassifiziert wird, tritt bei Wahl von $\lambda_{max}^{(j)}$ nicht auf. Wählt man jedoch $\lambda^{(j)}$ kleiner als $\lambda_{max}^{(j)}$, so wird die Wolke der Normalen nicht mehr komplett vom Parallelotop überdeckt. Somit gibt es Normale, die außerhalb des Parallelotops liegen und damit als behandlungsbedürftig klassifiziert werden ■

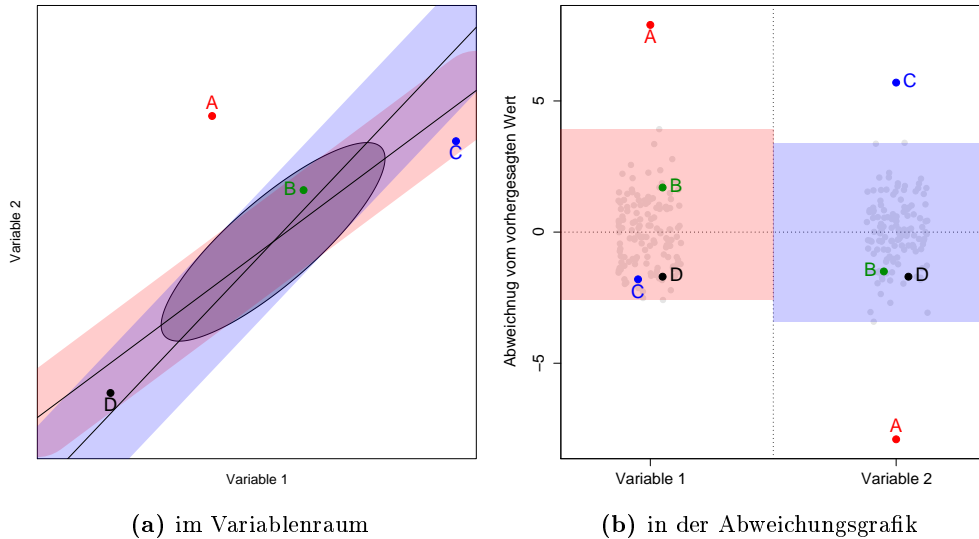


Abbildung 3.1.: Geometrie der Methode. (a) zeigt die Referenzgruppe, repräsentiert durch eine graue Normellipse. Es wurden die Regressionsgeraden sowie Bänder gemäß $\lambda_{max}^{(j)}$ eingezeichnet. Es werden vier verschiedene Punktkonfigurationen betrachtet. (b) zeigt die zugehörige Abweichungsgrafik mit den entsprechenden Bändern und Punkten.

3.3. Änderung der Abweichungen

Die Änderung einer Variablen hat Auswirkungen auf alle Abweichungen $r^{(1)}, \dots, r^{(k)}$ eines Individuums. Vergleicht man z.B. auf Seite 24 Abbildung 2.4a mit der Abweichungsgrafik 2.4c nach der Änderung von SNB, so stellt man fest, dass sich die Positionen aller Werte des Individuums verändert haben. Diese Feststellung verwundert nicht, da jede Variable in jeder Gleichung des Systems (2.28) vorkommt (Seite 18) - sei es als Response oder Prädiktor.

Unter der Annahme eines linearen Zusammenhangs der Form (2.6) interessiert die Auswirkung bei Änderung von Variable x_i auf die Abweichung $r^{(j)}$. Die Bestimmung der partiellen Ableitung liefert

$$\frac{\partial r^{(j)}}{\partial x_i} = \begin{cases} 1 & i = j \\ -\hat{\beta}_{i-1}^{(j)} & i > j \\ -\hat{\beta}_i^{(j)} & i < j. \end{cases} \quad (3.8)$$

Sämtliche partiellen Effekte lassen sich in einer $k \times k$ -Matrix \mathbf{A} , die ab jetzt die *Änderungsmatrix* genannt wird, zusammenfassen. Die Zeilen stellen die jeweiligen Variablen dar, deren Abweichungen man betrachtet. Die Spalten repräsentieren die zu ändernden Variablen.

$$\mathbf{A} := \begin{pmatrix} \frac{\partial r_1}{\partial x_1} & \dots & \frac{\partial r_1}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_k}{\partial x_1} & \dots & \frac{\partial r_k}{\partial x_k} \end{pmatrix} = \begin{pmatrix} 1 & -\hat{\beta}_1^{(1)} & \dots & \dots & -\hat{\beta}_{k-1}^{(1)} \\ -\hat{\beta}_1^{(2)} & 1 & -\hat{\beta}_2^{(2)} & \dots & -\hat{\beta}_{k-1}^{(2)} \\ -\hat{\beta}_1^{(3)} & -\hat{\beta}_2^{(3)} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & -\hat{\beta}_{k-1}^{(k-1)} \\ -\hat{\beta}_1^{(k)} & \dots & & -\hat{\beta}_{k-1}^{(k)} & 1 \end{pmatrix} \quad (3.9)$$

Es fällt auf, dass die Änderungsmatrix der Matrix im Satz 3.3 entspricht. Mit dieser Matrix lässt sich elegant beschreiben, was bei Anpassung mehrerer Variablen passiert. Es ist offensichtlich, dass die Änderungen von mehreren Variablen einfach additiv sind. Werden die einzelnen Änderungen, die an den Variablen x_1, \dots, x_k vorgenommen werden, in einem Vektor $v \in \mathbb{R}^k$ zusammengefasst, so ergibt sich als Gesamtänderung der Abweichungen der Vektor $Av =: w \in \mathbb{R}^k$.

Beispiel 3.6. Man betrachte Abweichungsgrafik 2.4d, in der SNB um 2 und ML-NSL um 4 Grad verändert wurden (die Ausgangssituation ist durch Abb. 2.4a gegeben). Als Vektor

der vorgenommenen Änderungen ergibt sich $v_0 := (0, 0, 0, -4, -2)^T$. Für den Datensatz und die Gleichungen (2.29) (Seite 19) erhält man folgende Änderungsmatrix

$$A_0 := \begin{pmatrix} 1.00 & -0.10 & 0.06 & -0.06 & -0.81 \\ -0.38 & 1.00 & -0.12 & -0.20 & 0.74 \\ 0.54 & -0.25 & 1.00 & -0.00 & 0.13 \\ -0.41 & -0.33 & -0.00 & 1.00 & 1.07 \\ -0.74 & 0.17 & 0.01 & 0.15 & 1.00 \end{pmatrix} \quad (3.10)$$

Aus $A_0 v_0 =: w_0$ resultieren die Gesamtänderungen

$$w_0 = (1.86, -0.68, -0.26, -6.14, -2.60)^T,$$

der Residuen für das Individuum, welche durch Vergleich der Grafik 2.4a mit Grafik 2.4d beobachtet werden können. ■

Wie muss man die Werte x_1, \dots, x_k ändern, damit man selbst vorgegebene Abweichungen $r' := (r'^{(1)}, \dots, r'^{(k)})$ erreicht? Hat das Individuum die Abweichungen $r := (r_1, \dots, r_k)$ und betrachtet man den Vektor w als die nötigen Änderungen der Abweichungen, so lässt sich das Problem folgendermaßen formulieren:

$$r + w = r' \quad (3.11)$$

Falls \mathbf{A} invertierbar ist, ergibt sich als eindeutige Lösung

$$v = A^{-1}(r' - r). \quad (3.12)$$

Beispiel 3.7. Angenommen, man hat Abweichungen $r_0 = (-3.4, -0.3, -6.1, 12.3, 4.6)^T$ (vgl. Grafik 2.4a) gegeben und möchte die Residuen $r'_0 = (-1, -1, -5.5, 6, 1.5)^T$ haben. Ist A_0 wie im Beispiel 3.6 gegeben, dann ergeben sich mit Hilfe von

$$A_0^{-1} = \begin{pmatrix} 3.70 & -0.54 & -0.33 & -0.48 & 3.95 \\ -2.25 & 1.87 & 0.40 & 0.87 & -4.19 \\ -3.03 & 0.88 & 1.33 & 0.58 & -3.90 \\ -3.09 & 1.40 & 0.41 & 1.95 & -5.68 \\ 3.61 & -0.94 & -0.38 & -0.80 & 5.53 \end{pmatrix}$$

die benötigten Änderungen der Variablen x_1, \dots, x_5 als

$A_0^{-1}(r'_0 - r_0) = (-0.2, 1, 1.3, -2.8, -3)^T$. Man betrachte die neue Position des Individuums in Grafik 3.2. Im unteren Bildrand sind die Änderungen eingetragen. Die geringfügigen Abweichungen von den theoretischen Werten kommen durch gerundete Darstellung zustande. ■

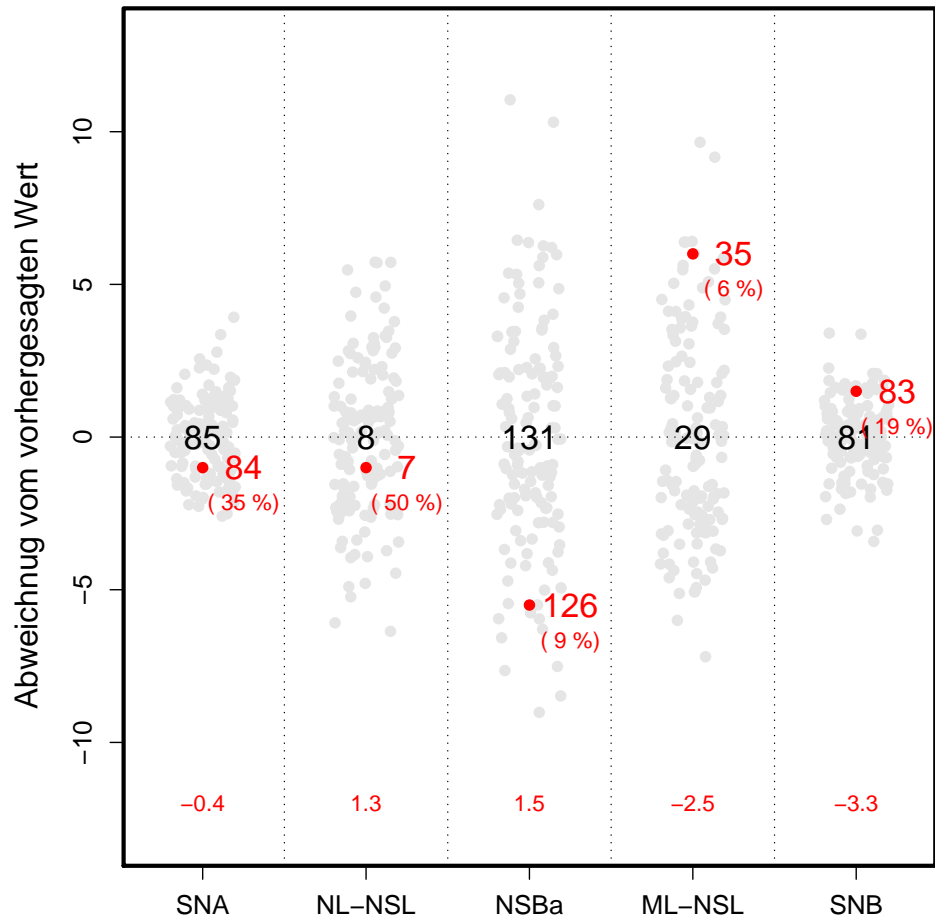


Abbildung 3.2.: Anwendung der Änderungsmatrix. Eine Abweichungsgrafik im gewohnten Stil. Die Zahlen am unteren Rand stellen die nötigen Änderungen dar, um zu dieser gewünschten Stellung der Winkel zu kommen.

Kapitel 4.

Die Harmoniebox nach Segner

In diesem Kapitel wird eine aus der Kieferorthopädie bekannte Methode vorgestellt - die *Harmoniebox* nach Segner. Nach einer kurzen Einführung in die Konstruktion der Box und ihre Anwendung folgt eine geometrische Interpretation und es werden einige Probleme der Harmoniebox genannt. Eine ausführliche Beschreibung dieser Methode findet sich in [SA98] und wird an dieser Stelle nur kurz umrissen.

4.1. Konstruktion der Box

Die Intention der Harmoniebox ist es, mit Hilfe eines Winkels (Segner wählt dafür SNA) die im Mittel zu erwartende Größe in der Referenzgruppe für die anderen kephalometrischen Variablen (NL-NSL, NSBa, ML-NSL und SNB) zu bestimmen. Die Harmoniebox wurde speziell für diese fünf Winkel entwickelt. Die Konstruktion erfolgt in folgenden Schritten:

1. Schritt: Es werden folgende Regressionsmodelle betrachtet:

$$\begin{aligned} \text{NL-NSL} &= \alpha_0^{(2)} + \alpha_1^{(2)} \text{SNA} + F^{(2)} \\ \text{NSBa} &= \alpha_0^{(3)} + \alpha_1^{(3)} \text{SNA} + F^{(3)} \\ \text{ML-NSL} &= \alpha_0^{(4)} + \alpha_1^{(4)} \text{SNA} + F^{(4)} \\ \text{SNB} &= \alpha_0^{(5)} + \alpha_1^{(5)} \text{SNA} + F^{(5)} \end{aligned} \tag{4.1}$$

2. Schritt: Angewendet auf den koreanischen Datensatz schätzt man mit der Referenzgruppe (Normale) die folgenden Regressionsgleichungen:

$$\begin{aligned}
 \text{NL-NSL} &= 60.21 - 0.61 \text{ SNA} + \mathbf{R}^{(2)} \\
 \text{NSBa} &= 200.26 - 0.82 \text{ SNA} + \mathbf{R}^{(3)} \\
 \text{ML-NSL} &= 100.71 - 0.84 \text{ SNA} + \mathbf{R}^{(4)} \\
 \text{SNB} &= -0.75 + 0.98 \text{ SNA} + \mathbf{R}^{(5)}
 \end{aligned} \tag{4.2}$$

In Tabelle 4.1 sind die Standardfehler, sowie die Signifikanzniveaus der Tests der Nullhypothese $H_0 : \alpha_i^{(j)} = 0$, für $i = 0, 1$ und $j = 2, \dots, 5$, für die einzelnen Parameter angeben.

Response	Intercept	SNA
NL-NSL	9.039 ***	0.111 ***
NSBa	10.833 ***	0.133 ***
ML-NSL	11.729 ***	0.144 ***
SNB	4.518	0.055 ***
Signifikanz level: <0.001 ***, 0.01 **, 0.05 *		

Tabelle 4.1.: Standardfehler und Signifikanzniveaus für die bivariate Regression. Alle Parameter, außer der Intercept bei der Vorhersage von SNB durch SNA, sind hoch signifikant von Null verschieden.

Doch welcher Anteil der Variabilität im Datensatz lässt sich durch die Regressionen erklären? In Tabelle 4.2 sind die Bestimmtheitsmaße, die korrigierten Bestimmtheitsmaße und die Standardfehler der Regression angegeben.

Response	\mathcal{R}^2	$\bar{\mathcal{R}}_c^2$	$\hat{\sigma}$
NL-NSL	0.19	0.18	3.21
NSBa	0.22	0.22	3.85
ML-NSL	0.20	0.20	4.17
SNB	0.70	0.70	1.61

Tabelle 4.2.: Bestimmtheitsmaße und Standardfehler der bivariaten Regression. \mathcal{R}^2 und $\bar{\mathcal{R}}_c^2$, welche den Erklärungsgehalt der Regressionen widerspiegeln, stimmen zum größten Teil überein und fallen relativ gering aus (abgesehen von SNB)

Es folgt eine visuelle Überprüfung der Regressionen und Residuen in Abbildung 4.1, welche eine Tabelle von diagnostischen Grafiken zeigt. Die Grafiken zeigen jedoch keine größeren Auffälligkeiten.

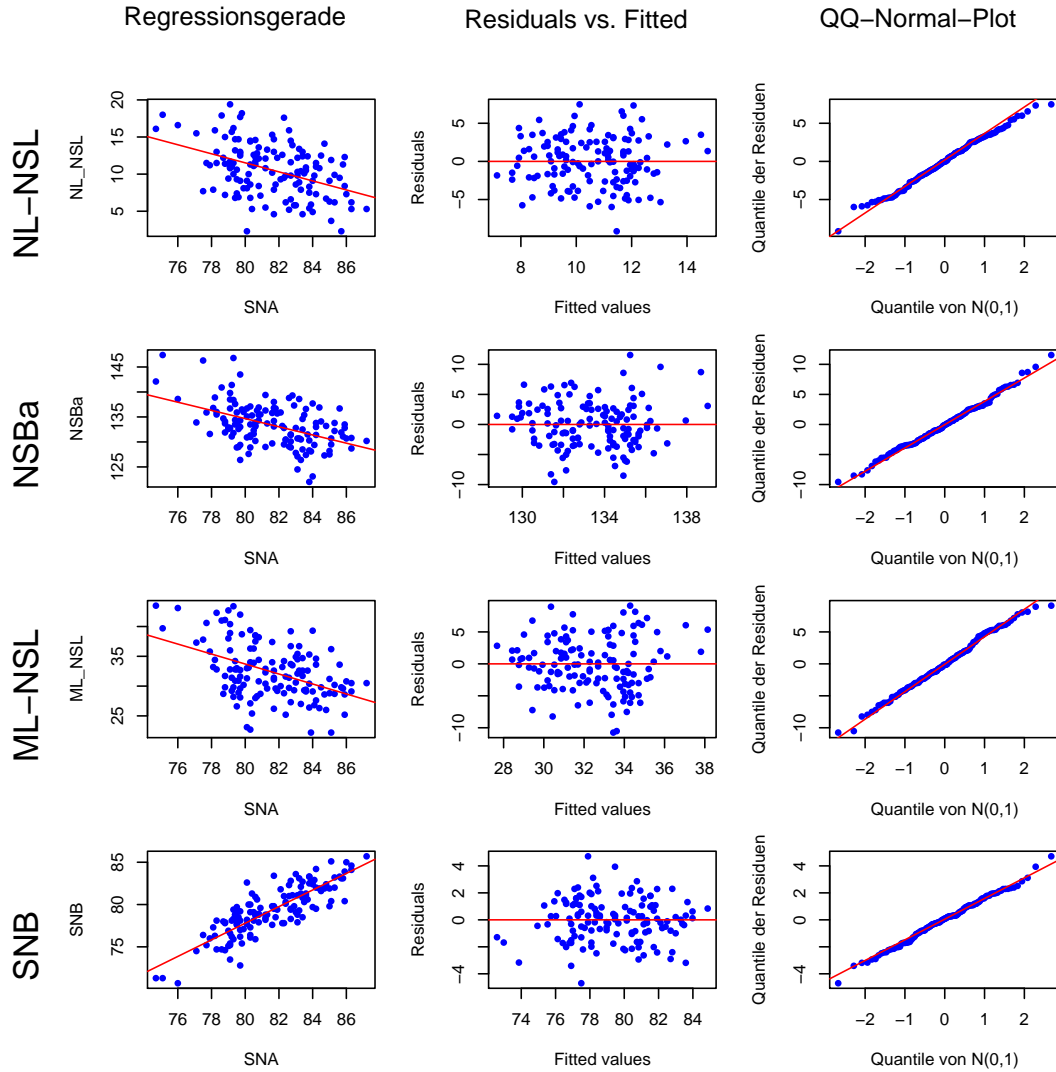


Abbildung 4.1.: Diagnostische Grafiken für die bivariate Regression. In der ersten Spalte befinden sich Streudiagramme mit der jeweiligen abhängigen Variablen (2. Achse) und SNA als Predictor (1. Achse) sowie die dazugehörigen Regressionsgeraden. Spalte zwei enthält die Residuen aufgetragen gegen die gefitteten (vorhergesagten) Werte. Die dritte Spalte überprüft mit Hilfe eines QQ-Plots die Annahme der Normalverteilung der Residuen.

3. Schritt: Für verschiedene Ausprägungen von SNA werden die dazugehörigen vorhergesagten Werte gemäß der Gleichungen (4.2) bestimmt und tabellarisch zusammengefasst. Die resultierende Tabelle wird die *Harmoniebox* genannt. Für den Datensatz ergibt sich die Darstellung in Abbildung 4.2a.

Zu sehen sind neben den fünf Winkeln auch die zwei zusätzlichen Winkel ANB und ML-NL. Diese sind als die Differenzen $ANB := SNA - SNB$ und $ML-NL := ML-NSL - NL-NSL$ definiert. Für die Harmoniebox hat das zur Folge, dass sie die zeilenweisen Differenzen aus den entsprechenden Spalten darstellen. Die beiden Winkel stehen folglich nur indirekt mit dem Modell (4.1) in Zusammenhang und sind für die weiteren Überlegungen nicht von Belang.

Es fällt die unterschiedliche Skalierung in den einzelnen Spalten auf. Wird die Tabelle von oben nach unten gelesen, so steigen die Werte von SNA und SNB, während die anderen fallen. Die Begründung für diese Tatsache liegt in den unterschiedlichen Vorzeichen der Steigungen der Regressionsgeraden (4.2).

4.2. Das Harmonieschema

4. Schritt: Harmonische Winkelkombinationen sind bei Segner definiert als alle Werte auf einer horizontalen Linie in der Harmoniebox. Diese wird *Harmonielinie* genannt. Jedoch wird ein gewisser Abweichungsbereich von der Linie für jeden Winkel (für jede Spalte in der Tabelle) zugelassen. Die erlaubte Abweichung nach oben und unten für Variable j entspricht dem Standardfehler $\hat{\sigma}^{(j)}$ der multivariaten Regressionen (siehe Tabelle 2.2 auf Seite 20).

Die Harmonielinie, inklusive der schwarzen Abweichungsgrenzen, vergleiche Abb. 4.2b, wird *Harmonieschema* genannt (des öfteren wird die Tabelle in Verbindung mit dem Schema auch Harmoniebox bezeichnet). Das Harmonieschema lässt sich nach belieben nach oben oder unten verschieben, wobei alle Punkte innerhalb des Schemas Winkelkombinationen darstellen, die man als normal klassifizieren würde und Punkte außerhalb als behandlungsbedürftig.

5. Schritt: Die Positionierung des Schemas für ein bestimmtes Individuum ist nicht beliebig. Segner gibt in [SA98] auf Seite 82 folgende Richtlinien an:

1. Die zentrale Linie muss exakt horizontal orientiert sein.

2. Die Summe der Abstände der einzelnen Meßwerte von der Zentrallinie oberhalb dieser Linie soll gleich derer unterhalb sein.
3. Dabei finden die Variablen SNA und SNB stärkere Berücksichtigung, indem diese Abstände dreifach zählen.
4. Liegen einzelne Meßwerte eindeutig weit außerhalb, so handelt es sich um "Ausreißer", die für die Ermittlung des Gesichtstyps keine Berücksichtigung finden sollten und deren Abstand deshalb nicht in die Rechnung unter Punkt 1 und 2 einbezogen wird.

Mit Zentrallinie ist die Harmonielinie gemeint. Punkt 2 entspricht einer Mittelwertsbestimmung, in die SNA und SNB dreifach eingehen (Punkt 3). Punkt 4 ist problematisch, da nicht genau festgelegt wird, wann ein Wert als Ausreißer zu betrachten ist. Ein weiteres Problem ist die Tatsache, dass man generell Ausreißer erkennen möchte und sie deshalb nicht von der Analyse ausschließen sollte. Dies ist einer der Gründe für die Entwicklung des MH Ansatzes.

Um die beiden Methoden später vergleichbar zu machen, wurde in dieser Arbeit folgende Implementierung gewählt: Von den auf die Spalte SNA horizontal projizierten Werten in der Tabelle wird der Median bestimmt (grobe Bestimmung des Gesichtstyps) und eine imaginäre Harmonielinie eingezeichnet. Von dieser Linie aus werden wieder Abweichungsgrenzen, die eine Länge von vier Standardfehlern der jeweiligen multivariaten Regression haben, nach oben und unten betrachtet. Alle Werte in diesen Grenzen werden für Punkt 2 und Punkt 3 herangezogen, um die einzutragende Harmonielinie zu bestimmen.

Offensichtlich wird durch die Wahl von vier Standardfehlern ein großer Bereich abgedeckt. Dies ist jedoch beabsichtigt, damit nur extreme Werte als Ausreißer betrachtet werden. Natürlich gibt es auch andere Möglichkeiten, Punkt 4 umzusetzen bzw. die Option Punkt 4 zu vernachlässigen. Das hier genannte Vorgehen wurde lediglich der Vollständigkeit halber und zur Rekonstruktion der Ergebnisse angegeben.

Beispiel 4.1. *Das Grundkonzept der Harmoniebox wird nochmals anhand des fiktiven Patienten aus Kapitel 2.4 erläutert, dessen Winkelwerte in Tabelle 2.3 auf Seite 22 zu finden sind. Trägt man diese in die Tabelle ein und bestimmt mittels der oberen Vorgehensweise die Position des Harmonieschemas, so ergibt sich das in Abbildung 4.2b dargestellte Bild. ■*

Diese durch die Harmoniebox dargestellte einfache Herangehensweise bildet einen möglichen Grundstein für die weitere Behandlungsplanung.

ANB	SNA	NL-NSL	NSBa	ML-NSL	SNB	ML-NL
2	60	24			58	27
	62	23	150	49	60	
	64	22	148	47	62	26
	66	21			64	
	68	20	146	45	66	25
	70	19	144	43	68	
	72	18			70	
	74	17	142	41	72	24
	76	16	140	39	74	
	78	15			76	
	80	14	138	37	78	23
	82	13	136	35	80	
	84	12			82	
	86	11	134	33	84	22
2.5	88	10	132	31	86	21
	90	9			88	
	92	8	130	29	90	20
	94	7			92	
	96	6	128	27	94	19
	98	5	126	25	96	
	100	4				
		3	124	23		
		2	122	21		
		1	120	19		18
		0				
		-1	118	17		

(a)

ANB	SNA	NL-NSL	NSBa	ML-NSL	SNB	ML-NL
2	60	24			58	27
	62	23	150	49	60	
	64	22	148	47	62	26
	66	21			64	
	68	20	146	45	66	25
	70	19	144	43	68	
	72	18			70	
	74	17	142	41	72	24
	76	16	140	39	74	
	78	15			76	
	80	14	138	37	78	23
	82	13	136	35	80	
	84	12			82	
	86	11	134	33	84	22
2.5	88	10	132	31	86	21
	90	9			88	
	92	8	130	29	90	20
	94	7			92	
	96	6	128	27	94	19
	98	5	126	25	96	
	100	4				
		3	124	23		
		2	122	21		
		1	120	19		18
		0				
		-1	118	17		

(b)

Abbildung 4.2.: Die Harmoniebox. (a) Zeigt die Box mit den zusätzlichen Winkeln ANB und NL-NSL. (b) In der Harmoniebox sind die Winkelwerte des fiktiven Patienten eingezeichnet sowie Harmonieline und Harmonieschema positioniert. Alle Werte außer NL-NSL liegen außerhalb des Schemas. Das Individuum ist somit behandlungsbedürftig.

4.3. Eigenschaften

4.3.1. Probleme der Harmoniebox

Bei der Konstruktion der Harmoniebox fallen folgende Fragen und Probleme auf:

1. Obwohl aus Tabelle A.1 in Anhang A.2.2 (Seite 62) ersichtlich ist, dass zwischen allen Winkeln deutliche Korrelationen bestehen, werden durch die Betrachtung der bivariaten Gleichungen (4.1) nur die Paare von Variablen betrachtet, in denen SNA enthalten ist und somit die restlichen und wichtigen mehrdimensionale Zusammenhänge vernachlässigt.
2. Für das Harmonieschema werden die Standardfehler der multivarianten Regression verwendet, obwohl die Tabelle aus den bivariaten Gleichungen entsteht. Außerdem fällt das Harmonieschema durch die Wahl nur eines Standardfehlers relativ klein aus, was zur Konsequenz hat, dass viele ursprünglich normale mit der Harmoniebox nicht mehr als normal sondern als behandlungsbedürftig eingeschätzt werden.
3. Ein weiteres Problem der Harmoniebox ist, dass mittlerweile die verschiedensten Modifikationen der Methode aufgetreten sind. Franchi, Baccetti und McNamara benutzen in [FBM98] zum Beispiel SNB anstelle von SNA als unabhängige Variable, was zu anderen Regressionsmodellen führt, was wiederum in einer anderen Tabelle/Harmoniebox resultiert. Die Problematik wird an folgendem Beispiel verdeutlicht:

Beispiel 4.2. *In Abbildung 4.3 wird in der oberen Grafik eine Harmoniebox mit SNA als unabhängiger Variable dargestellt und in der unteren eine Box mit SNB (man tausche in (4.1) SNA mit SNB). Beide Tabellen wurden auf Basis des vorliegenden koreanischen Datensatzes erstellt. Je weiter man sich vom Mittelwert (Schwerpunkt der Daten) entfernt, desto mehr unterscheiden sich die Tabellen. ■*

Es gibt noch weitere Modifikationen. Ein anderer Vorschlag zur Positionierung des Harmonieschemas, vorgestellt durch Tollaro, Baccetti und Franchi in [TBF96], besagt, dass die Werte des Individuums in der Tabelle auf eine Spalte der Tabelle zu projizieren sind. Von diesen neuen projizierten Werten wird der Mittelwert bestimmt und an diesem die Harmonielinie orientiert. Offensichtlich führt dies zu unterschiedlichen Diagnosen in der Patientenbewertung. Um diese Uneindeutigkeit zu umgehen, beschränkt sich diese Arbeit auf die Vorgaben von Segner.

ANB	SNA	NL-NSL	NSBa	ML-NSL	SNB	ML-NL
	60	24			58	27
	62	23	150	49	60	
	64	22			62	26
	66	21	148	47	64	
	68	20	146	45	66	25
	70	19	144	43	68	
	72	18	142	41	70	24
	74	17	140	39	72	
	76	16	138	37	74	23
	78	15	136	35	76	
	80	14	134	33	78	22
	82	13			80	
	84	12	132	31	82	21
	86	11	130	29	84	
	88	10	128	27	86	20
	90	9	126	25	88	
	92	8	124	23	90	19
	94	7	122	21	92	
	96	6	120	19	94	18
	98	5	118	17	96	
	100	4			98	
		3			100	

(a) SNA als unabhängige Variable

ANB	SNA	NL-NSL	NSBa	ML-NSL	SNB	ML-NL
8	68	25	147		60	27
7	70	23	145	51	62	
6	72	21	143	49	64	26
	74	19	141	47	66	
5	76	17	139	45	68	25
	78	15	137	43	70	
4	80	13	135	41	72	24
	82	11	133	39	74	
3	84	9	131	37	76	23
	86	7	129	35	78	
2	88	5	127	33	80	22
	90	3	125	31	82	
1	92	1	123	29	84	21
	94	0	121	27	86	
0	96	-1	119	25	88	20
	98	-3		23	90	
-1	100	-5		21	92	19
				19	94	
-2				17	96	18
				15	98	
-3				13	100	17
				11		
-4						16

(b) SNB als unabhängige Variable

Abbildung 4.3.: Verschiedene Harmonieboxen. (a) zeigt eine Harmoniebox, die mit SNA als unabhängiger Variable konstruiert worden ist. (b) zeigt eine Box mit SNB als unabhängige Variable. In beide Tabellen ist der Datenmittelwert (untere Linie) markiert. Die obere Linie in beiden Harmonieboxen ist an einem SNB Wert von 62 ausgerichtet, um die Unterschiede in beiden Tabellen zu verdeutlichen.

4.3.2. Geometrische Interpretation

Es wird der 5-dimensionale Variablenraum betrachtet, wobei die Referenzgruppe der Normalen als Punktwolke in diesem Raum liegt. Bei der bivariaten Regression werden per Definition nur zwei Variablen betrachtet. Die Punktwolke wird somit in den durch die zwei betrachteten Variablen (SNA und die jeweilige Zweite) dargestellten 2-dimensionalen Raum (eine Ebene) projiziert. Dies wird für die vier Paare von Variablen entsprechend den Gleichungen (4.1) auf Seite 35 ausgeführt und die dazugehörigen Regressionsgeraden in die jeweilige Ebene eingezeichnet. Diese Projektionen entsprechen den Streudiagrammen aus der linken Spalte in Abb. 4.1.

Beispiel 4.3. In Abbildung 4.4a ist der Fall $k = 3$ für die Winkel SNA, SNB und ML-NSL gezeigt. Die schwarzen Punkte stellen die einzelnen Normalen, die hellen die Projektionen in die entsprechende Ebene dar. Anhand der eingezeichneten bivariaten Regressionsgeraden kann man die Streudiagramme aus Abbildung 4.1 wiedererkennen. ■

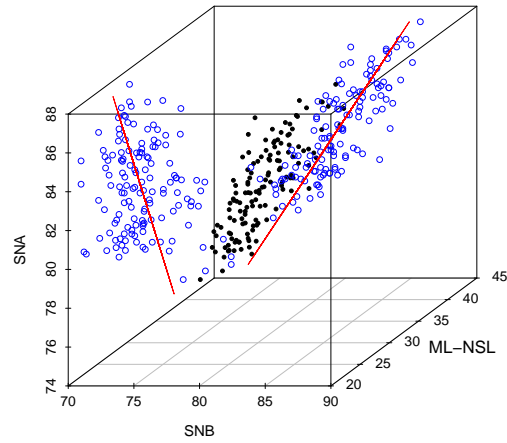
Da man sich aber in einem 5-dimensionalen Raum befindet, werden aus diesen "Geraden" 4-dimensionale Hyperebenen (die restlichen drei Variablen können beliebige Werte annehmen). Es ist mittlerweile bekannt, dass alle bivariaten Regressionsgeraden durch den Schwerpunkt der beiden Variablen gehen, woraus folgt, dass die Hyperebenen durch den Gesamtschwerpunkt verlaufen. Da man gemäß (4.1) vier Regressionsgleichungen hat, gibt es insgesamt vier Hyperebenen. Es wird nun gezeigt, dass sich diese in einer Geraden schneiden:

Satz 4.4. Die durch das System (4.1) bestimmten 4-dimensionalen Hyperebenen

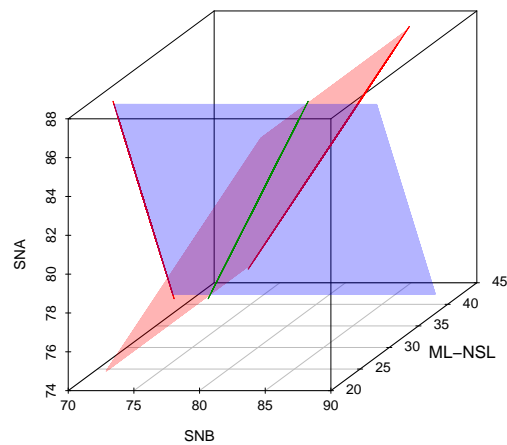
$$\begin{aligned} \{(x_1, \dots, x_5) \in \mathbb{R}^5 \mid x_2 &= \hat{\alpha}_0^{(2)} + \hat{\alpha}_1^{(2)} x_1\} \\ \{(x_1, \dots, x_5) \in \mathbb{R}^5 \mid x_3 &= \hat{\alpha}_0^{(3)} + \hat{\alpha}_1^{(3)} x_1\} \\ \{(x_1, \dots, x_5) \in \mathbb{R}^5 \mid x_4 &= \hat{\alpha}_0^{(4)} + \hat{\alpha}_1^{(4)} x_1\} \\ \{(x_1, \dots, x_5) \in \mathbb{R}^5 \mid x_5 &= \hat{\alpha}_0^{(5)} + \hat{\alpha}_1^{(5)} x_1\} \end{aligned}$$

schneiden sich im \mathbb{R}^5 in einer Geraden.

Beweis. Wie im Beweis von Satz 3.3 betrachte man die Hyperebenen in ihrer geordneten Koordinatendarstellung:



(a)



(b)

Abbildung 4.4.: *Der 3-dimensionale Variablenraum. (a) zeigt die schwarzen Datenpunkte im dreidimensionalen Raum. Die hellen Punkte stellen Projektionen in die zweidimensionale Ebene dar. In diesen sind die Regressionsgeraden eingezeichnet. (b) zeigt die durch die Geraden gehenden Hyperebenen welche sich wieder in einer Gerade schneiden*

$$\begin{aligned}\hat{\alpha}_0^{(2)} &= -\hat{\alpha}_1^{(2)}x_1 + x_2 \\ \hat{\alpha}_0^{(3)} &= -\hat{\alpha}_1^{(3)}x_1 + x_3 \\ \hat{\alpha}_0^{(4)} &= -\hat{\alpha}_1^{(4)}x_1 + x_4 \\ \hat{\alpha}_0^{(5)} &= -\hat{\alpha}_1^{(5)}x_1 + x_5\end{aligned}$$

Dieses inhomogene Gleichungssystem lässt sich schreiben als

$$\begin{pmatrix} -\hat{\alpha}_1^{(2)} & 1 & 0 & 0 & 0 \\ -\hat{\alpha}_1^{(3)} & 0 & 1 & 0 & 0 \\ -\hat{\alpha}_1^{(4)} & 0 & 0 & 1 & 0 \\ -\hat{\alpha}_1^{(5)} & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} \alpha_0^{(2)} \\ \alpha_0^{(3)} \\ \alpha_0^{(4)} \\ \alpha_0^{(5)} \end{pmatrix}$$

Die Matrix hat offensichtlich Rang 4. Die Dimension des Lösungsraums ist somit $5-4 = 1$, was einer Geraden entspricht. \square

Beispiel 4.5. Sei $k = 3$. Regressiert man zwei Variablen jeweils auf die dritte, wie z.B.

$$\begin{aligned}\text{NL-NSL} &= \alpha_0^{(2)} + \alpha_1^{(2)}\text{SNA} + F^{(2)} \\ \text{SNB} &= \alpha_0^{(5)} + \alpha_1^{(5)}\text{SNA} + F^{(5)}\end{aligned}$$

dann entsprechen die dazugehörigen Regressionsgeraden im 3-dimensionalen Raum 2-dimensionale Ebenen. Diese Ebenen schneiden sich in einer Geraden. Man vergleiche dazu die beiden Grafiken in Abbildung 4.4 ■

Die Harmoniebox wird durch die Schnittgerade der vier Hyperebenen repräsentiert. Jede horizontale Linie in der Tabelle stellt somit einen Punkt auf dieser Schnittgeraden dar. Betrachtet man eine feste Position der Harmonielinie in der Tabelle, so würde man einen Punkt auf der Schnittgeraden betrachten und einen 5-dimensionalen Quader, mit Seitenlängen dem Standardfehler $\hat{\sigma}^{(j)}$ der multivariaten Regressionen (2.28) entsprechend, um diesen Punkt legen. Lässt man das Harmonieschema die Harmoniebox von oben nach unten durchwandern, so stelle man sich einen Schlauch vor, der die Schnittgerade umhüllt.

Die Positionierung des Schemas lässt sich wie folgt erklären: Das Individuum, das einen Punkt im 5-dimensionalen Raum darstellt, wird in Richtung jeder Variable auf diese

Schnittgerade projiziert. Der Punkt hat somit fünf Projektionen auf diese Gerade. Betrachtet man diese fünf Werte eindimensional (also auf der Gerade), so wird von ihnen ein gewichtetes Mittel genommen, was der Positionierung der Harmonielinie entspricht. Um diesen Punkt wird der im vorherigen Absatz beschriebene 5-dimensionale Quader mit Seitenlängen dem Standardfehler $\hat{\sigma}^{(j)}$ der multivariaten Regressionen entsprechend eingezeichnet.

Beispiel 4.6. Eine schematische Darstellung, wobei die Normalen wieder durch eine Ellipse aus der bivariaten Normalverteilung dargestellt werden, befindet sich in Abbildung 4.5. Es wird deutlich, dass aufgrund der Verwendung lediglich eines einfachen Standardfehlers die Referenzwolke nur teilweise überdeckt wird und dass der durch den Schlauch definierte Bereich auch über die Wolke hinausgeht. Betrachtet man das Individuum P , so werden dessen Winkelwerte in jeder Variablen und auch in Richtung der Variablen auf die Schnittgerade projiziert. Aus den entstehenden Punkten wird das gewichtete Mittel genommen, welches die Position des Schema darstellt. ■

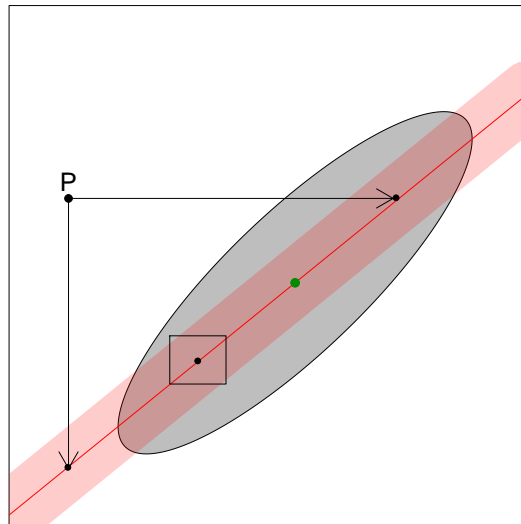


Abbildung 4.5.: Geometrische Skizze der Harmoniebox. Gezeigt wird ein Individuum P , eine die Normwolke repräsentierende Ellipse und die Schnittgerade, welche die Tabelle der Harmoniebox repräsentiert. Das Individuum wird in Variablenrichtung auf die Gerade projiziert und aus diesen Werten ein gewichtetes Mittel bestimmt. Dies entspricht der Position der Harmonielinie. Um diesen Punkt ist ein Rechteck mit Seitenlängen die dem Standardfehler der multivariaten Regression entsprechen eingezeichnet. Dies stellt das Harmonieschema dar.

Kapitel 5.

Vergleich von Harmoniebox und MH

In diesem Kapitel wird der MH Ansatz mit der *Harmoniebox* nach Segner verglichen. Im weiteren Verlauf werden die beiden Ansätze in Bezug auf die Regressionsgleichungen und die Klassifikationsgüte beim koreanischen Datensatz im Sinne der Sensitivität und Spezifität einander gegenübergestellt.

Schließlich wird beim Einsetzen der Behandlungsvorschläge aus der Harmoniebox in das MH Modell ein wichtiger Zusammenhang zwischen den beiden Modellen aufgedeckt und diskutiert. Dieser wird im letzten Abschnitt des Kapitels formal in einem allgemeinen Satz gezeigt.

5.1. Vergleich

Der offensichtliche Hauptunterschied zwischen der Box und dem MH Ansatz liegt darin, dass die Harmoniebox bivariate Regressionsmodelle mit einer festen Variable als Regressor verwendet und der MH Ansatz multivariate Gleichungen behandelt, wobei jede Variable als Response und alle anderen als Regressoren genommen werden.

Es werden zunächst die Regressionsgleichungen untersucht und die jeweiligen Bestimmtheitsmaße und Standardfehler gegenübergestellt. Im Anschluss wird die Klassifikationsgüte beider Methoden in Bezug auf den koreanischen Datensatz verglichen.

5.1.1. Regressionen

Multivariate Regressionsmodelle sind offensichtlich besser dazu geeignet mehrdimensionale Zusammenhänge zu erfassen als bivariate. Doch welchen Nutzen hat man von der Verwendung der multivariaten Gleichungen?

In Tabelle 5.1 sind das Bestimmtheitsmaß \mathcal{R}^2 und der Standardfehler der Regression $\hat{\sigma}$ für die bivariaten und multivariaten Gleichungen angegeben. Wenn man von NSBa absieht, haben sich alle Werte durch die Hinzunahme von drei weiteren Prädiktoren deutlich verbessert. NSBa scheint sich generell nicht gut durch die anderen vorliegenden kephalometrischen Variablen vorhersagen zu lassen. Zusammenfassend scheinen die multivariaten Gleichungen besser geeignet, die Zusammenhänge zwischen den Winkeln zu beschreiben.

Response	Bestimmtheitsmaße		Standardfehler	
	multi. Reg.	biv. Reg.	multi. Reg.	biv. Reg.
SNA	0.74	— — —	1.31	— — —
NL-NSL	0.47	0.19	2.61	3.21
NSBa	0.27	0.22	3.78	3.85
ML-NSL	0.48	0.20	3.40	4.17
SNB	0.82	0.70	1.26	1.61

Tabelle 5.1.: Vergleich der multivariaten mit der bivariaten Regression. Gezeigt sind das Bestimmtheitsmaß \mathcal{R}^2 und der Standardfehler der Regression $\hat{\sigma}$ für die bivariaten und multivariaten Gleichungen. Auf die Angabe des korrigierten Bestimmtheitsmaßes wurde verzichtet, da es in der multivariaten wie auch bivariaten Regression in etwa \mathcal{R}^2 entspricht.

5.1.2. Klassifikation von Individuen

Ohne kieferorthopädisches Fachwissen ist nicht bekannt, wie die Patienten im Datensatz behandelt werden sollten. Die einzige bekannte Information ist, dass ein Beobachter sie als behandlungsbedürftig bestimmt hat. Somit lassen sich die beiden Ansätze nur in Bezug auf die Klassifizierung (die Diagnose) der Individuen vergleichen.

Im Sinne der Sensitivität und Spezifität interessieren die beiden folgenden Fragen:

- Wie viel Prozent der Normalen werden von der jeweiligen Methode auch als Normale erkannt? (Spezifität)
- Wie viel Prozent der Patienten werden von der jeweiligen Methode als behandlungsbedürftig klassifiziert? (Sensitivität)

In Tabelle 5.2 ist der prozentuale Anteil der richtigen Klassifikation für beide Ansätze und beide Gruppen angegeben. Dabei wird die Konvention verwendet, dass sobald mindestens ein Wert außerhalb des Harmonieschemas liegt, die Person als behandlungsbedürftig einzustufen ist. Eine Vergrößerung des Harmonieschemas hat zur Konsequenz, dass mehr Normale richtig erkannt werden, geht aber mit einer Reduzierung der Patientenerkennung einher.

	Harmoniebox	MH Ansatz
Normale	25.4%	100%
Patienten	100%	98.8%

Tabelle 5.2.: Klassifikation der Individuen im jeweiligen Ansatz. Die Harmoniebox erkennt alle Patienten als solche. Jedoch werden aufgrund der Wahl von nur einem Standardfehler für das Harmonieschema lediglich 25.4% der Normalen erkannt. Der MH Ansatz hingegen erkennt mit der Verwendung von $\lambda_{max}^{(j)}$ alle Normalen und 98.8% der Patienten.

Tabelle 5.3 zeigt die Klassifikationsgüte der Harmoniebox bei Verwendung von zwei Standardfehlern. Es werden mehr Normale richtig klassifiziert, jedoch nicht mehr alle Patienten als behandlungsbedürftig erkannt.

	Harmoniebox	MH Ansatz
Normale	85.8%	83.6%
Patienten	90.7%	100%

Tabelle 5.3.: Klassifikation: Harmoniebox mit 2 Standardfehlern und MH mit $\lambda_{\sigma}^{(j)}(2)$. Die Klassifikationsrate beim MH Ansatz liegt für die Normalen bei 83.6%. Dafür werden jedoch alle Patienten auch als diese erkannt.

Zusammenfassend kann man sagen, dass die MH Methode mit $\lambda_{max}^{(j)}$ bezogen auf den koreanischen Datensatz, eine sehr gute Klassifikationsrate hat.

5.2. Zusammenhang zwischen den Ansätzen

Der MH Ansatz und die Harmoniebox gehen von einer unterschiedlichen Grundidee bzw. Fragestellung aus. Überspitzt formuliert:

MH Ansatz: Was erwartet man im Mittel für eine Variable, gegeben alle anderen Winkel?

Harmoniebox: Gegeben SNA, was erwartet man im Mittel für alle anderen Winkel?

Im Folgenden wird deutlich, dass es trotz dieser und weiterer offensichtlicher Unterschiede es einen Zusammenhang zwischen beiden Ansätzen gibt.

5.2.1. Harmonielinien im MH Ansatz

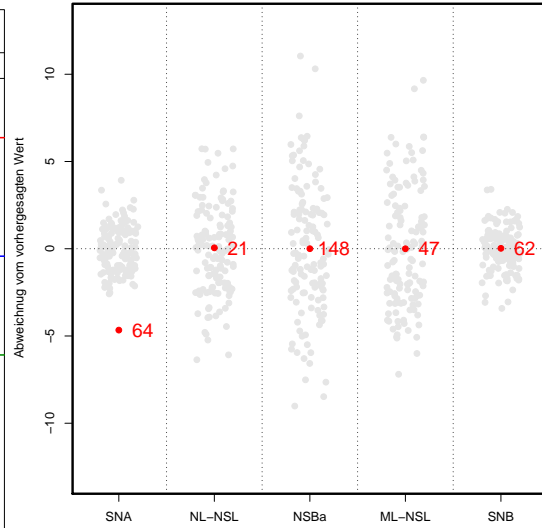
Die MH Methode macht im Gegensatz zur Harmoniebox keine wirklich konkreten Vorschläge, wie der Patient aussehen sollte. Sie zeigt dem Anwender lediglich die Position des Individuums im Vergleich zu den Normalen an und lässt ihn neue denkbare Winkelkombinationen überprüfen. Doch wie sehen die Harmonielinien der Harmoniebox im MH Ansatz aus? In Abbildung 5.1 werden dafür drei Fälle betrachtet.

Es fällt auf, dass alle Werte, außer denen von SNA, genau den Vorhersagen entsprechen und somit immer eine Abweichung von Null haben. Dies hat eine interessante geometrische Interpretation. Man stelle sich die Tabelle wieder als eine Gerade im 5-dimensionalen Raum vor, die durch die Normwolke geht (vgl. Grafik 4.4 S. 44). Die drei in der Tabelle eingezeichneten Harmonielinien entsprechen Punkten auf dieser Gerade. Nach Kapitel 3.1 gilt, dass jede Spalte in der Abweichungsgrafik des MH Ansatzes die Abweichung von einer Hyperebene darstellt. Bei der oberen Linie beobachtet man zum Beispiel in Grafik 5.1b, dass NL-NSL eine Abweichung von Null hat. Der Punkt liegt somit in der Hyperebene die NL-NSL als Response hat. Dies beobachtet man jedoch auch bei allen anderen Abweichungen, außer bei SNA. Der Punkt ist somit auch in diesen Hyperebenen enthalten. Der Punkt ist also im Schnitt dieser vier Hyperebenen enthalten.

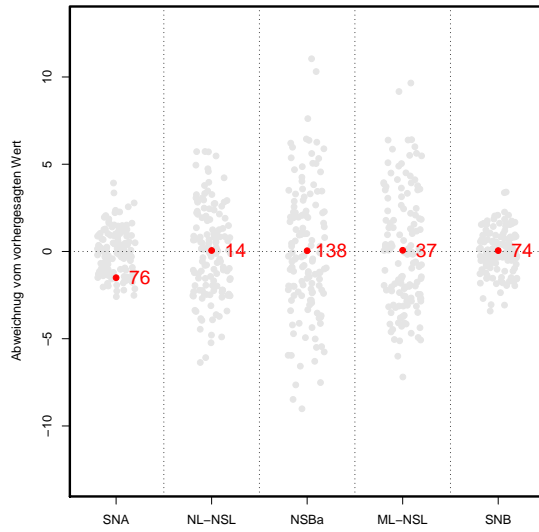
Diese Beobachtung kann man für jede mögliche Harmonielinie der Harmoniebox machen (Satz 4.4 S. 43). Die Tabelle, die bekanntlich die Schnittgerade aus vier Hyperebenen darstellt, ist somit mit dem Schnitt der vier Hyperebenen aus der multivariaten Regression (Korollar 3.4 S. 28) identisch.

	ANB	SNA	NL-NSL	NSBa	ML-NSL	SNB	ML-NL
2		60	24			58	27
		62	23	150	49	60	
		64	22	148	47	62	26
		66	21	146	45	64	
		68	20	144	43	66	25
		70	19	142	41	68	
		72	18	140	39	70	24
		74	17	138	37	72	
		76	16	136	35	74	23
		78	15	134	33	76	
2.5		80	14	132	31	80	22
		82	13	130	29	82	
		84	12	128	27	84	21
		86	11	126	25	86	
		88	10	124	23	88	20
		90	9	122	21	90	
		92	8	120	19	92	19
		94	7	118	17	94	
		96	6			96	18
		98	5				
		100	4				
			3				
			2				
			1				
			0				
			-1				

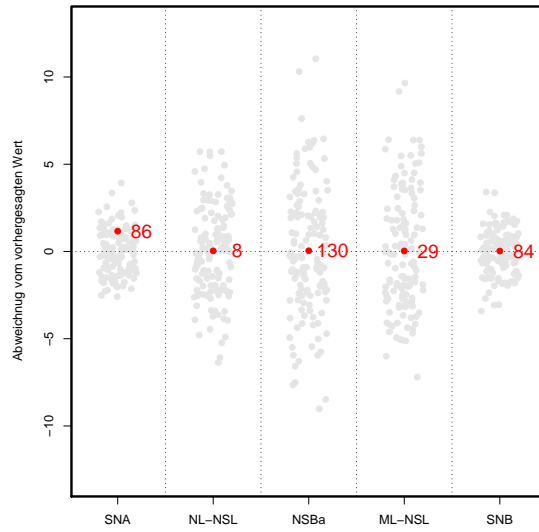
(a) Drei verschiedene Harmonielinien



(b) Die obere Harmonielinie



(c) Die mittlere Harmonielinie



(d) Die untere Harmonielinie

Abbildung 5.1.: Verschiedene Harmonielinien im MH Modell. (a) Harmoniebox, in der drei Harmonielinien markiert sind. (b)-(d) zeigt die jeweiligen Harmonielinien in der dazugehörigen Abweichungsgrafik. Man bemerkt, dass alle Variablen, außer SNA, eine Abweichung von Null haben.

Im folgenden Abschnitt wird die hier gemachte Beobachtung in einem Satz aufgestellt und bewiesen. Aus dem Satz wird folgen, dass man diese Beobachtung auch macht, wenn z.B. die Box in Abhängigkeit von SNB konstruiert wird. Die daraus resultierende Gerade würde wieder mit der Schnittgeraden der vier Regressionshyperebenen (wenn man die von SNB weglässt) aus dem MH Ansatz übereinstimmen. Dies liefert die Begründung für die Namenswahl der Methode. Denn alle möglichen Harmonieboxen, die durch die Wahl verschiedener Prädiktoren entstehen, sind sozusagen im MH Ansatz enthalten.

5.2.2. Mathematischer Zusammenhang

In diesem Abschnitt wird der n -dimensionale Raum der Beobachtungen und die Lage der n Normalen darin betrachtet. Jedes Merkmal $\mathbf{X}_j = (x_{1j}, \dots, x_{nj})^T$, $j = 1, \dots, k$, stellt einen n -elementigen Vektor in diesem Raum dar. $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$ ist der *Einsvektor*.

Definition 5.1. Die Notation $\mathcal{P}_{\mathbf{X}}\mathbf{Y}$ bezeichne die orthogonale Projektion des Vektors $\mathbf{Y} \in \mathbb{R}^n$ auf den durch $\mathbf{X} \in \mathbb{R}^n$ aufgespannten Unterraum $\{c\mathbf{X} | c \in \mathbb{R}\}$. Analog bezeichnet $\mathbf{X}_1 \oplus \dots \oplus \mathbf{X}_l$ den durch $\mathbf{X}_1, \dots, \mathbf{X}_l$ aufgespannten Unterraum und $\mathcal{P}_{(\mathbf{X}_1 \oplus \dots \oplus \mathbf{X}_l)}\mathbf{Y}$ ist die orthogonale Projektion von \mathbf{Y} auf diesen.

Für alle $V_1, V_2, W \in \mathbb{R}^n$ und $a \in \mathbb{R}$ gelten die folgenden Eigenschaften in Bezug auf die orthogonale Projektion:

1. $\mathcal{P}_W W = W$
2. $\mathcal{P}_W(V_1 + V_2) = \mathcal{P}_W V_1 + \mathcal{P}_W V_2$
3. $\mathcal{P}_W(aV_1) = a\mathcal{P}_W V_1$
4. $\mathcal{P}_{V_1}(\mathcal{P}_{(V_1 \oplus V_2)}(W)) = \mathcal{P}_{V_1} W$

Bei einer multivariaten linearen Regressionsgleichung $\mathbf{Y} = \hat{\beta}_0 \mathbf{1} + \hat{\beta}_1 \mathbf{X}_1 + \dots + \hat{\beta}_k \mathbf{X}_k + \mathbf{F}$ liegt $\hat{\mathbf{Y}} := \hat{\beta}_0 \mathbf{1} + \hat{\beta}_1 \mathbf{X}_1 + \dots + \hat{\beta}_k \mathbf{X}_k$ in einem Unterraum der Form

$$\mathcal{M} := \mathbf{1} \oplus \mathbf{X}_1 \oplus \dots \oplus \mathbf{X}_k = \{\hat{\beta}_0 \mathbf{1} + \hat{\beta}_1 \mathbf{X}_1 + \dots + \hat{\beta}_k \mathbf{X}_k | \hat{\beta}_0, \dots, \hat{\beta}_k \in \mathbb{R}\}$$

Minimiert man die Residuen mittels der Methode der kleinsten Quadrate, so gilt $\hat{\mathbf{Y}} = \mathcal{P}_{\mathcal{M}}\mathbf{Y}$, weil die orthogonale Projektion das Abstandsquadrat minimiert. Man überlege

sich, dass im Spezialfall, in dem man das Regressionsmodell $\mathbf{Y} = \alpha\mathbf{X} + \mathbf{F}$ betrachtet $\hat{\mathbf{Y}} = \mathcal{P}_{\mathbf{X}}\mathbf{Y}$ gilt.

Im Weiteren wird von zentrierten Daten ausgegangen. Dies hat den Verlust des konstanten Terms in den Regressionsgleichungen zur Folge. Da sich an den restlichen Parametern der Regression nichts ändert, bleiben die gewonnenen Aussagen auch für den Fall mit konstantem Term gültig. Der Einfachheit halber werden die zentrierten Daten nicht anders bezeichnet. Es sei daran erinnert, dass $\mathbf{X}^{(j)} := (\mathbf{X}_1, \dots, \mathbf{X}_{j-1}, \mathbf{X}_{j+1}, \dots, \mathbf{X}_k)$. Für den folgenden Satz definiert man zusätzlich $\beta^{(j)} := (\beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_k)^T$ und $F^{(j)}$ als den Fehler der einfachen linearen Regression mit \mathbf{X}_j als Zielvariable. Nach dieser Vorarbeit lässt sich der Satz beweisen:

Satz 5.2. *Seien $\mathbf{X}_1, \dots, \mathbf{X}_k \in \mathbb{R}^n$ linear unabhängige, zentrierte Datenvektoren. Sei $i \in \{1, \dots, k\}$ beliebig gewählt und man betrachte alle einfachen linearen Regressionsgleichungen der Form $\mathbf{X}_l = \alpha^{(l)}\mathbf{X}_i + F^{(l)}$ mit $l \neq i$. Außerdem habe man eine multivariate lineare Regressionsgleichung der Form $\mathbf{X}_j = \mathbf{X}^{(j)}\beta^{(j)} + \mathbf{F}^{(j)}$ mit $j \neq i$ gegeben. Sei o.B.d.A. $i = 1$ und $j = k$.*

Schätzt man die Parameter $\alpha^{(l)}$, $l = 2, \dots, k$ und $\beta^{(j)} = (\beta_1, \dots, \beta_{k-1})^T$ mit der Methode der kleinsten Quadrate, so ergibt sich $\hat{\mathbf{X}}_2 := \hat{\alpha}^{(2)}\mathbf{X}_1, \dots, \hat{\mathbf{X}}_k := \hat{\alpha}^{(k)}\mathbf{X}_1$ und $\tilde{\mathbf{X}}_k := \hat{\beta}_1\mathbf{X}_1 + \dots + \hat{\beta}_{k-1}\mathbf{X}_{k-1}$. Dann gilt:

$$\hat{\beta}_1\mathbf{X}_1 + \hat{\beta}_2\hat{\mathbf{X}}_2 + \dots + \hat{\beta}_{k-1}\hat{\mathbf{X}}_{k-1} = \tilde{\mathbf{X}}_k.$$

Beweis. Da man nur zentrierte Daten betrachtet, ist

$$\begin{aligned} \hat{\mathbf{X}}_2 &= \mathcal{P}_{\mathbf{X}_1}\mathbf{X}_2 \\ &\vdots \\ \hat{\mathbf{X}}_k &= \mathcal{P}_{\mathbf{X}_1}\mathbf{X}_k. \end{aligned} \tag{5.1}$$

Desweiteren gilt:

$$\tilde{\mathbf{X}}_k = \mathcal{P}_{(\mathbf{X}_1 \oplus \dots \oplus \mathbf{X}_{k-1})}\mathbf{X}_k = \sum_{j=1}^{k-1} \hat{\beta}_j\mathbf{X}_j \tag{5.2}$$

Die Behauptung ergibt sich unter Verwendung der weiter oben genannten Eigenschaften der orthogonalen Projektion wie folgt:

$$\hat{\mathbf{X}}_k = \mathcal{P}_{\mathbf{X}_1} \mathbf{X}_k = \mathcal{P}_{\mathbf{X}_1} (\mathcal{P}_{(\mathbf{X}_1 \oplus \dots \oplus \mathbf{X}_{k-1})} \mathbf{X}_k) \quad (5.3)$$

$$= \mathcal{P}_{\mathbf{X}_1} \left(\sum_{j=1}^{k-1} \hat{\beta}_j \mathbf{X}_j \right) \quad (5.4)$$

$$= \mathcal{P}_{\mathbf{X}_1} (\hat{\beta}_1 \mathbf{X}_1) + \sum_{j=2}^{k-1} \mathcal{P}_{\mathbf{X}_1} (\hat{\beta}_j \mathbf{X}_j) \quad (5.5)$$

$$= \hat{\beta}_1 \mathbf{X}_1 + \sum_{j=2}^{k-1} \hat{\beta}_j \hat{\mathbf{X}}_j \quad (5.6)$$

□

Dieser Satz stellt die mathematische Erklärung des beobachteten Phänomens aus Abschnitt 5.2.1 dar. Er besagt, dass, wenn man die Vorhersagen der einfachen linearen Regression wie sie bei Segner vorkommen, in den MH Ansatz einsetzt, so stimmen diese mit den multivariaten Vorhersagen überein. Die Ausnahme ist die multivariate Gleichung, die den Regressor der bivariaten Regressionen als Zielvariable hat.

Dies erkennt man daran, wenn man \mathbf{X}_1 mit SNA identifiziert und \mathbf{X}_k mit SNB. Dann besagt der Satz: Setzt man die Werte einer Harmoniebox in die Regressionsgleichung ein, welche SNB als Response hat, dann entspricht die Vorhersage der multivariaten Gleichung derer der bivariaten. Somit hat man ein Residuum von Null. Der Satz sagt weiterhin, dass dies für alle multivariaten Gleichungen, außer der mit SNA als Zielvariablen, gilt (dies entspricht der Bedingung "Außerdem habe man eine multivariate lineare Regressionsgleichung der Form $\mathbf{X}_j = \mathbf{X}^{(j)} \beta^{(j)} + \mathbf{F}^{(j)}$ mit $j \neq i$ " aus Satz 5.2). Man betrachte dazu nochmals Abb. 5.1.

Doch der Satz geht noch einen Schritt weiter. Er gilt für jedes zu Beginn fest gewählte \mathbf{X}_i . Das bedeutet, dass eine analoge Aussage für jede mögliche konstruierbare Harmoniebox gültig ist. Zum Beispiel könnte man $\mathbf{X}_i = \mathbf{X}_5$ setzen und dies mit SNB identifizieren. Der Satz macht dann eine analoge Aussage für alle Harmonielinien aus Tabelle 4.3b auf Seite 42.

Der MH Ansatz beinhaltet somit in gewisser Weise sämtliche Harmonieboxen in sich. Dadurch ist es ihm möglich eine größere Vielfalt von harmonischen Gesichtern darzustellen.

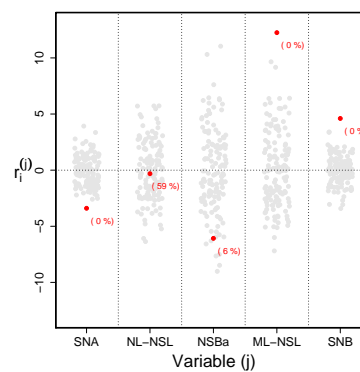
Kapitel 6.

Diskussion

Der MH Ansatz wurde entwickelt zur Beschreibung der als harmonisch klassifizierten Winkelkombinationen in einer Norm Population. Ausgehend von Variablen X_1, \dots, X_k , basiert er auf der Betrachtung der multivariaten linearen Regressionsgleichungen

$$\begin{aligned} \mathbf{X}_1 &= f_1(\mathbf{X}^{(1)}) + \mathbf{F}^{(1)} \\ &\vdots \\ \mathbf{X}_k &= f_k(\mathbf{X}^{(k)}) + \mathbf{F}^{(k)}, \end{aligned}$$

wobei $\mathbf{X}^{(j)} := (\mathbf{X}_1, \dots, \mathbf{X}_{j-1}, \mathbf{X}_{j+1}, \dots, \mathbf{X}_k)$, welche durch eine Referenzgruppe von als harmonisch klassifizierten Individuen geschätzt werden. Mittels der resultierenden Residuen lassen sich nicht klassifizierte Individuen mit Normalen vergleichen. Zur besseren Visualisierbarkeit wurde die Abweichungsgrafik eingeführt. Klassifiziert werden Individuen durch ein Kriterium $\lambda^{(j)}$, $j = 1, \dots, k$, was die maximal tolerierten Abweichungen pro Variable darstellt.



Der in dieser Arbeit vorgestellte MH Ansatz soll Kieferorthopäden in die Lage versetzen, sich unter Berücksichtigung der mehrdimensionalen Zusammenhänge anschaulich einen Überblick über die relative Position eines Individuums zu verschaffen. Der Ansatz wurde ursprünglich für die fünf Winkel aus dem koreanischen Datensatz konstruiert, lässt sich aber auch auf weitere kephalometrische Variablen anwenden.

Die allgemein formulierte Regel $\lambda^{(j)}$ birgt die Möglichkeit der Individualisierung der Klassifikationskriterien durch den Anwender. Während zum Beispiel $\lambda_{max}^{(j)}$ alle Abweichungen aus der Referenzgruppe zulässt, erlauben andere Regeln mehr Steuerung durch den Kieferorthopäden.

Die vorgestellten Schritte aus Kapitel 2.2 zur Konstruktion des Ansatzes ermöglichen ohne Weiteres eine Umsetzung der Methode am Computer. Die gezeigten Abweichungsgrafiken entstammen einer Implementierung mit dem Statistikpaket R. Eine Einführung in R erhält man mittels [Ver04] und insbesondere für die Regressionsanalyse mit [Far04].

Da die Definition eines Normalen nicht eindeutig ist, hängt die Klassifikation eines Individuums von der zugrunde liegenden Referenzgruppe ab. In unserem Fall handelt es sich um Individuen mit Angle Kl.I, die zusätzlich aufgrund ihres harmonischen Äußeren ausgewählt wurden. Es spricht jedoch nichts dagegen auch harmonische Angle Kl.III oder Kl.II Individuen (die in diesem Sinne auch "normal" sind) in die Referenzgruppe aufzunehmen.

In unserem Fall ist es weitestgehend möglich die Winkel SNA, NL-NSL, NSBa, ML-MSL und SNB unabhängig voneinander zu verändern. Die Annahme der Konstanz aller anderen Winkel beim Anpassen einer Variable ist bei Wahl beliebiger anderer Winkel in der Regel jedoch nicht erfüllt. Aufgrund des skelettalen Zusammenhangs haben Änderungen einer Variable oftmals Auswirkungen auf andere.

Die durch Segner vorgestellte Harmoniebox beschreibt harmonische Winkelkombinationen in Abhängigkeit einer Variablen. Da die Box speziell für die Winkel SNA, NL-NSL, NSBa, ML-NSL und SNB geschaffen wurde, werden folgende Regressionsgleichungen betrachtet

$$\begin{aligned} \text{NL-NSL} &= \alpha_0^{(2)} + \alpha_1^{(2)} \text{SNA} + F^{(2)} \\ \text{NSBa} &= \alpha_0^{(3)} + \alpha_1^{(3)} \text{SNA} + F^{(3)} \\ \text{ML-NSL} &= \alpha_0^{(4)} + \alpha_1^{(4)} \text{SNA} + F^{(4)} \\ \text{SNB} &= \alpha_0^{(5)} + \alpha_1^{(5)} \text{SNA} + F^{(5)} \end{aligned}$$

Diese werden für verschiedene Ausprägungen von SNA tabellarisch zusammengefasst. Harmonische Winkelkombinationen stellen nach Segner alle Werte auf einer horizontalen

Linie in dieser Tabelle dar. Für die Klassifikation eines Individuums wird die Linie mittels einer bestimmten Vorgehensweise positioniert, wobei Abweichungen von jeweils einem Standardfehler der entsprechenden multivariaten Regression zugelassen werden.

	ANB	SNA	NL-NSL	NSBa	ML-NSL	SNB	ML-NL
2		60	24			58	27
		62	23	150	49	60	
		64	22	148	47	62	26
		66	21	146	45	64	
		68	20	144	43	66	25
		70	19	142	41	68	
		72	18	140	39	70	24
		74	17	138	37	72	
		76	16	136	35	74	23
		78	15	134	33	76	
		80	14	132	31	78	22
		82	13	130	29	80	
		84	12	128	27	82	21
		86	11	126	25	84	
		88	10	124	23	86	20
2.5		90	9	122	21	88	19
		92	8	120	19	90	
		94	7	118	17	92	18
		96	6			94	
		98	5			96	
		100	4				
			3				

Problematisch an der Harmoniebox ist neben der Vernachlässigung der multivariaten Zusammenhänge die Frage nach der unabhängigen Variablen. Die Verwendung eines anderen Merkmals als Prädiktor liefert eine andere Regressionsgerade und folglich eine andere Harmoniebox. Somit ist es möglich, dass für ein Individuum in verschiedenen Boxen unterschiedliche Diagnosen gestellt werden.

Die Wahl lediglich eines Standardfehlers hat zur Folge, dass nicht alle Normale als diese erkannt werden. Die hohe Sensitivität der Box geht somit sehr zu Lasten ihrer Spezifität. Dies lässt die Harmoniebox im direkten Vergleich nicht so gut abschneiden wie den MH Ansatz. In unserem Fall erzielt $\lambda_{max}^{(j)}$ sehr gute Ergebnisse.

Die hier vorgestellte Methode bietet ein flexibleres Harmoniekonzept als die Harmoniebox und berücksichtigt eine größere Vielfalt von als harmonisch wahrgenommenen Gesichtern. Die harmonischen Kombinationen aus der Box sind auch im MH Ansatz harmonisch, jedoch ist die Spezifität bei der MH Methode unter Verwendung von $\lambda_{max}^{(j)}$ bei 100% und geht nicht zu Lasten der Patientenerkennung.

Es sei nochmals betont, dass keine Handlungsempfehlungen für Kieferorthopäden gegeben werden, sondern eine Darstellung der momentanen Situation des Individuums visualisiert wird, was der weiteren Behandlungsplanung dient. Es ist jedoch möglich denkbare Vorgehensweisen für eine eventuell nötige Behandlung zu überprüfen. Auch wenn die Kieferorthopädie sich in Zukunft weiter entwickelt und neue Apparaturen und Modelle auf den Markt kommen, so stellt der MH Ansatz eine Möglichkeit dar, mehrdimensionale und komplexe Zusammenhänge anschaulich und anwenderfreundlich zu betrachten.

Anhang A.

Der Datensatz

Die Daten wurden von Dr. Yong-min Jo an der Yonsei Universität Seoul/Südkorea für Promotionszwecke gesammelt. Sie beinhalten 220 koreanische Individuen, im Alter von 18 bis 24 Jahren, mit insgesamt über 70 Variablen (Strecken und Längen). Dem zugänglichen Datensatz liegen jedoch außer der Patientennummer nur 6 Variablen zugrunde.

A.1. Die Variablen

Klasse Angle Klasse I (Normale) und Angle Klasse III (Patienten).

Die relative Lagebeziehung des menschlichen Ober- und Unterkiefers wird weltweit nach Edward H. Angle, siehe [Ang13], durch die so genannten Angle-Klassen beschrieben. 134 Individuen stammen aus Angle Klasse I (vgl. Abbildung A.1a). Diese Personen, welche zusätzlich wegen ihres harmonischen Äußeren ausgewählt wurden, stellen die Normalen und somit die Referenzgruppe dar. Des Weiteren enthält der Datensatz 86 Patienten, welche Angle Klasse III entstammen (vgl. Abb. A.1b).

SNA (Abb. A.2a): Dieser Winkel entspricht dem Prognathiegrad des Oberkiefers. Er ist Ausdruck der anterior-posterioren Position des A-Punktes in Relation zur anterioren Schädelbasis in Form von NSL (Nasion-Sella-Linie).

SNB (Abb. A.2c): Dieser Winkel entspricht dem Prognathiegrad des Unterkiefers in Relation zur Schädelbasislinie. Da die Patienten der Klasse III entstammen, weisen sie definitionsgemäß einen hohen SNB Winkel auf.

NL-NSL (Abbildung A.2c): Neigungswinkel des Oberkiefers. Zur Bestimmung des Winkels wird die Schädelbasislinie (NSL) in Richtung der Nasallinie (NL) verschoben.

ML-NSL (Abbildung A.2d): Neigungswinkel des Unterkiefers. Zur Bestimmung des Winkels wird NSL in Richtung ML verschoben.

NSBa (Abbildung A.2e): Schädelbasiswinkel. Beschreibt die Abknickung der hinteren Schädelbasis.

Sämtliche Winkel liegen in der Einheit "Grad" vor.

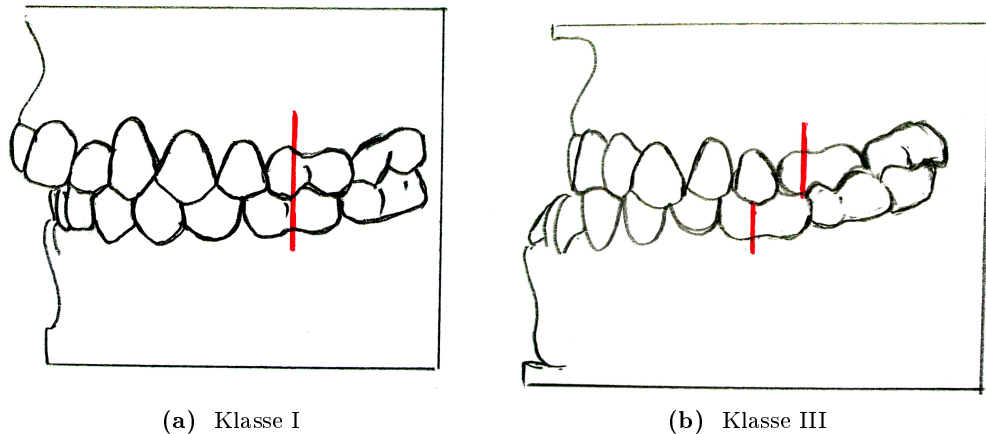


Abbildung A.1.: Angle Klassen. (a) zeigt Angle Klasse I, die dadurch definiert ist, dass der vordere zur Wange zeigende Höcker des oberen ersten großen Backenzahns zwischen die beiden großen buccalen Höcker des unteren ersten Backenzahns zeigt. (b) zeigt Angle Klasse III. Diese ist dadurch charakterisiert, dass der vordere zur Wange zeigende Höcker des oberen ersten Backenzahns hinter die Trennlinie der beiden großen buccalen Höcker des unteren ersten Molaren zeigt. Quelle: Dr. Georg Risse, Institute of Biofunctional Orthodontics

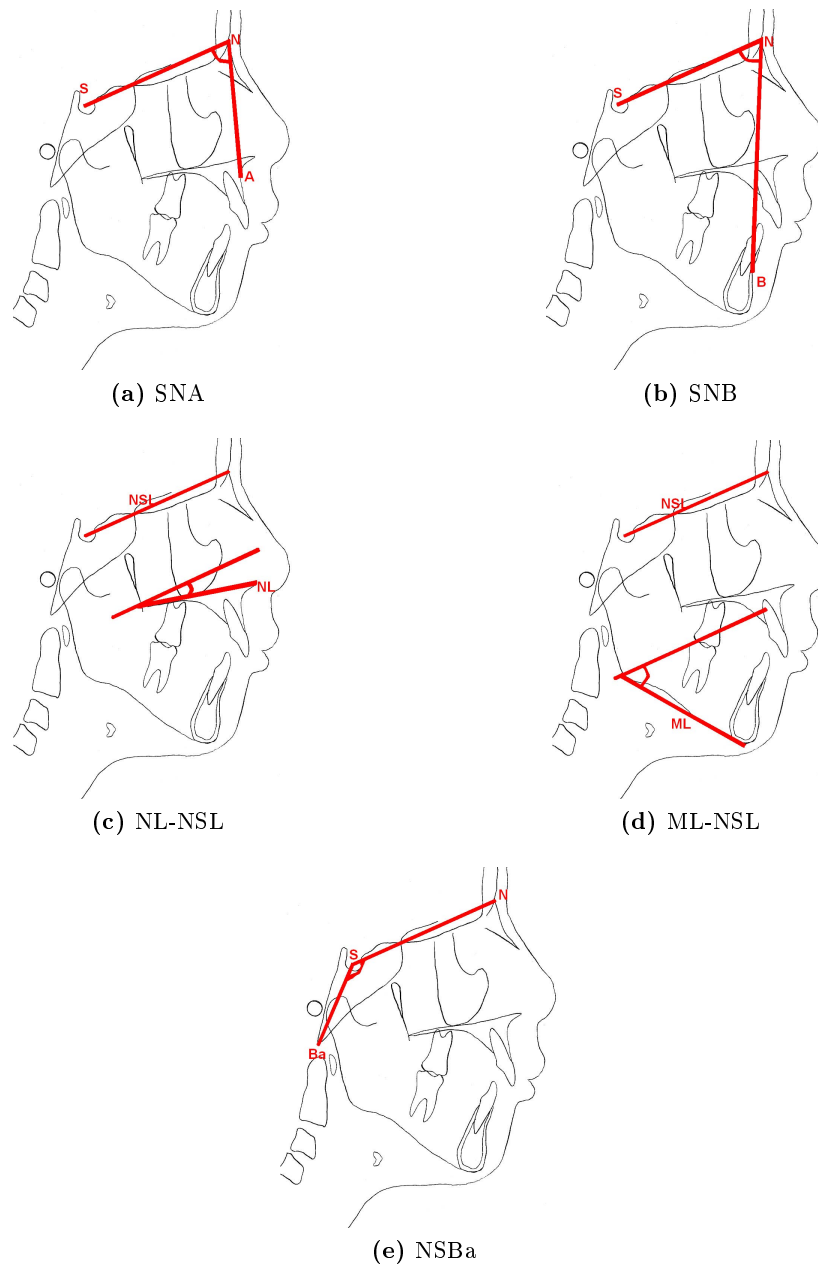


Abbildung A.2.: Darstellung der fünf Winkel. Die Winkel werden in der sagittalen Gesichtsebene gezeigt. Bei den Winkeln NL-NSL und ML-NSL wird die NSL Linie, die an der oberen Schädelbasis liegt, parallel verschoben, um diese zu bestimmen.

A.2. Deskriptive Statistiken

A.2.1. Verteilungen

Die Verteilung der Winkel in beiden Gruppen wird in Abbildung A.3 in der ersten Spalte mittels Dichtekurven verglichen. Es fallen die deutlichen Unterschiede bei SNB auf, bei dem die Patienten größere Winkelwerte erzielen. In der zweiten (Normale) und dritten Spalte (Patienten) sind QQ-Plots, um die Verteilung der Variablen mit der Normalverteilung zu vergleichen. An den Rändern fallen geringfügige Abweichungen auf.

Abbildung A.4 zeigt neben Mittelwert und Standardabweichung die einzelnen Beobachtungen für alle Winkel (Angaben in Grad) in beiden Gruppen. Die Patientengruppe weist in allen Variablen eine größere Streuung auf als die Normgruppe. Des Weiteren fällt die große Diskrepanz im SNB Winkel auf. Die Mittelwerte wurden für jeden Winkel einzeln in beiden Gruppen mittels eines t-Tests verglichen und waren auf einem Niveau von $p < 0.01$ signifikant verschieden (ausgenommen NL-NSL, was nicht signifikant war).

A.2.2. Korrelationen in der Normgruppe

Da insbesondere harmonische Winkelkombination - die Verhältnisse, in denen die Winkel zueinander stehen - in der Normgruppe von Interesse sind, sei ein besonderes Augenmerk auf die Korrelation der Variablen gelegt (siehe Tabelle A.1). Man kann somit von annähernd linearen Zusammenhängen ausgehen, welche nochmals in Abbildung A.5 mittels Streudiagrammen dargestellt werden.

	NL-NSL	NSBa	ML-NSL	SNB
SNA	-0.43	-0.47	-0.45	0.84
NL-NSL		0.39	0.58	-0.62
NSBa			0.32	-0.48
ML-NSL				-0.65

Tabelle A.1.: Korrelationen innerhalb der Normgruppe. Zwischen allen Variablen bestehen eindeutige Korrelationen, die aufgrund der Stichprobengröße ($n = 134$) hoch signifikant von Null verschieden sind ($p < 0.001$).

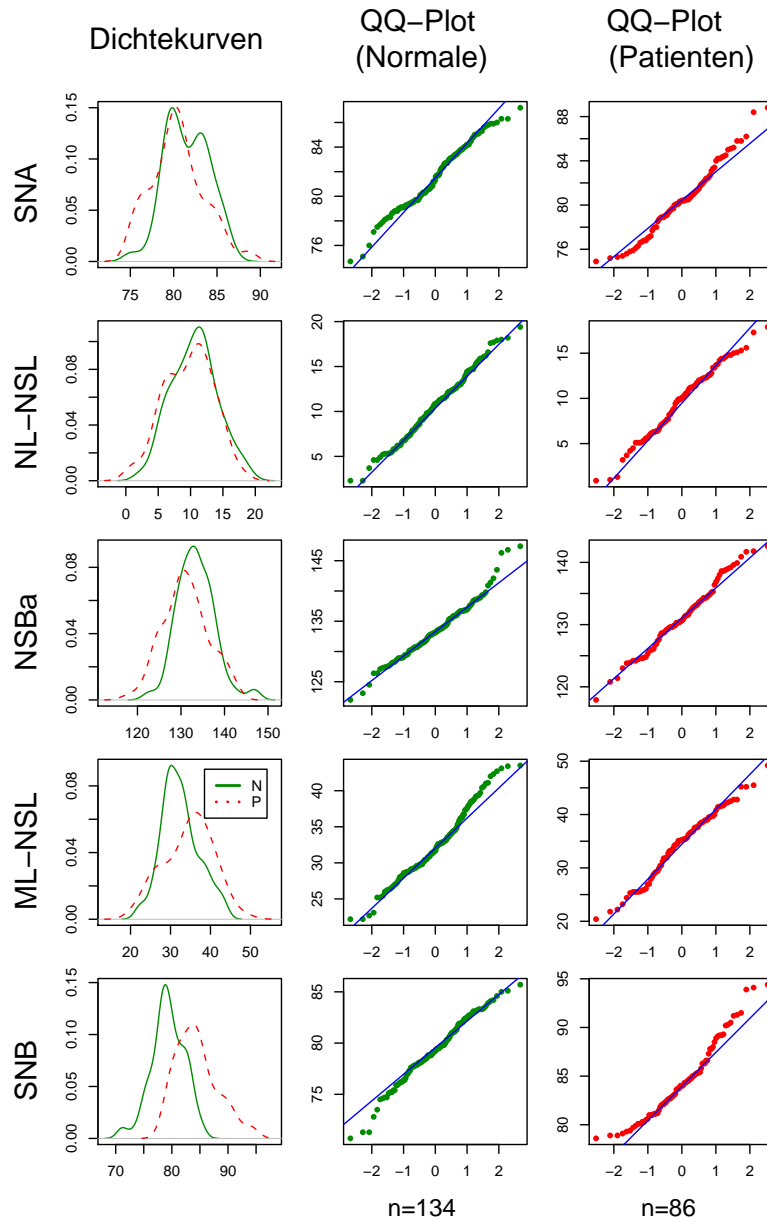


Abbildung A.3.: Verteilung der Winkel in Norm- und Patientengruppe. In jeder Zeile dieser Tabelle befindet sich ein Winkel. Die erste Spalte stellt die Verteilung des entsprechenden Winkels in Norm- und Patientengruppe mittels Dichtekurven gegenüber. Die zweite Spalte vergleicht mittels eines QQ-Plots die Verteilung der Normalen mit der Normalverteilung. Die dritte Spalte zeigt dies für die Patienten.

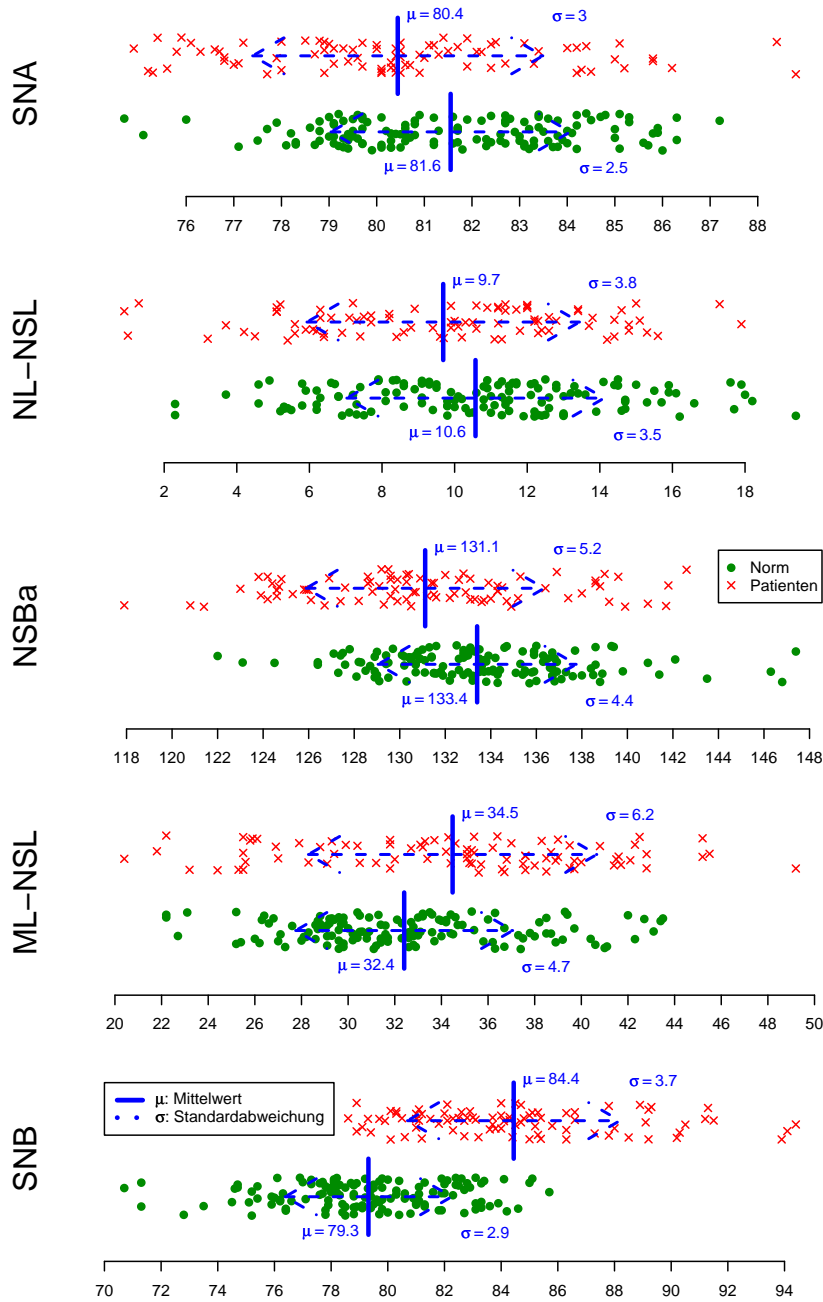


Abbildung A.4.: Winkel in Norm- und Patientengruppe für alle Individuen. Für jeden Winkel ist jedes Individuum als Punkt (Normale) oder Kreuz (Patienten) eingezeichnet. Der Mittelwert und die Standardfehler sind eingetragen.

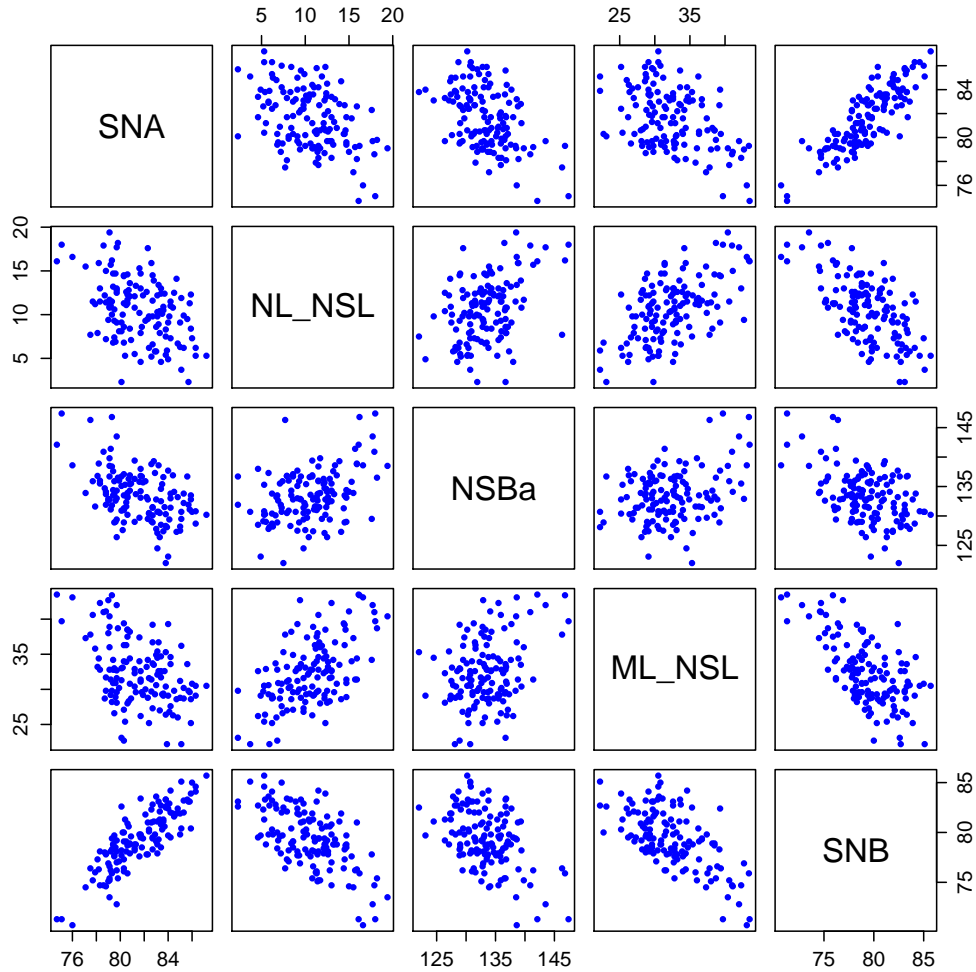


Abbildung A.5.: Streudiagramme der fünf Variablen. Für alle Paarungen von Variablen sind Streudiagramme dargestellt, um die möglichen Korrelationen und linearen Zusammenhänge zu verdeutlichen.

Literaturverzeichnis

- [Ang13] E. H. Angle. *Die Okklusionsanomalien der Zähne*. Meusser Verlag, Berlin, 1913.
- [BAH90] S. E. Bishara, E. M. Abdalla, and B. J. Hoppens. Cephalometric comparisons of dentofacial parameters between egyptian and north american adolescents. *Am J Orthod Dentofacial Orthop.*, 56:413–421, 1990.
- [Beu06] A. Beutelspacher. *Lineare Algebra*. Vieweg, 2006.
- [BFGJF86] S. E. Bishara, A. Fernandez Garcia, J. R. Jakobsen, and J. A. Fahl. Mesiodistal crown dimensions in mexico and the united states. *The Angle Orthodontist*, 56:315–323, 1986.
- [DS98] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley Series in Probability and Statistics, Berlin, 1998.
- [Far04] J. J. Faraway. *Linear Models with R*. Chapman & Hall/CRC, 2004.
- [FBM98] L. Franchi, T. Baccetti, and J. A. McNamara. Cephalometric floating norms for north american adults. *The Angle Orthodontist*, 68(6):497–502, 1998.
- [FKPT04] L. Fahrmeir, R. Künstler, I. Pigeot, and G. Tutz. *Statistik - Der Weg zur Datenanalyse*. Springer, 2004.
- [Han02] A. Handl. *Multivariate Analysemethoden. Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS*. Springer, Berlin, 2002.
- [ID90] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *VIS '90: Proceedings of the 1st conference on Visualization '90*, pages 361–378, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

- [Jär86] S. Järvinen. Floating norms for the anb angle as guidance for clinical considerations. *Am J Orthod Dentofacial Orthop.*, 90:383–387, 1986.
- [Kle06] A. Klenke. *Wahrscheinlichkeitstheorie*. Springer Verlag, 2006.
- [Mah] L. Mahaini. *Kraniofaziale Strukturen syrischer und deutscher Probanden*. PhD thesis, Ludwig-Maximilians-Universität München.
- [Mah36] P. C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Science*, 12:49–55, 1936.
- [Ric66] A. Richardson. An investigation into the reproducibility of some points, places and lines used in cephalometric analysis. *American Journal Orthodont*, 52:637–651, 1966.
- [Rya97] T. P. Ryan. *Modern Regression Methods*. Wiley Series in Probability and Statistics, New York, 1997.
- [SA98] D. Segner and Hasund A. *Individualisierte Kephalmetrie*. Dr. D. Segner Verlag, Hamburg, 1998.
- [Seg89] D. Segner. Floating norms as a means to describe individual skeletal patterns. *European Journal of Orthodontics*, 11:214–220, 1989.
- [SN] M. A. Sevilla-Naranjilla. *Schädelmorphologie philippinischer und deutscher Probanden mit Angle-Klasse-1-Okklusion: Eine kephalometrische Studie*. PhD thesis, Ludwig-Maximilians-Universität München.
- [Sol66] B. Solow. The pattern of craniofacial associations. a morphological and methodological correlation and factor analysis study on young adults. *Acta Odontologica Scandinavia*, Supplement 46, 1966.
- [Ste53] C. Steiner. Cephalometrics for you and me. *American Journal Orthodont.*, 39:729–785, 1953.
- [TBF96] I. Tollaro, T. Baccetti, and L. Franchi. Floating norms for the assessment of craniofacial pattern in the deciduous dentition. *European Journal of Orthodontics*, 18:359–365, 1996.
- [Ver04] J. Verzani. *Using R for Introductory Statistics*. Chapman & Hall/CRC, 2004.

Abbildungsverzeichnis

1.1. Ein einführendes Beispiel zur Patientenbewertung	6
2.1. Die Abweichungsgrafik	13
2.2. Das Modell für $p = 2$	17
2.3. Diagnostische Grafiken für die multivariate Regression	21
2.4. Das Modell in Anwendung	24
3.1. Geometrie der Methode	30
3.2. Anwendung der Änderungsmatrix	33
4.1. Diagnostische Grafiken für die bivariate Regression	37
4.2. Die Harmoniebox	40
4.3. Verschiedene Harmonieboxen	42
4.4. Der 3-dimensionale Variablenraum	44
4.5. Geometrische Skizze der Harmoniebox	46
5.1. Verschiedene Harmonielinien im MH Modell	51
A.1. Angle Klassen	60
A.2. Darstellung der fünf Winkel	61
A.3. Verteilung der Winkel in Norm- und Patientengruppe	63
A.4. Winkel in Norm- und Patientengruppe für alle Individuen	64
A.5. Streudiagramme der fünf Variablen	65

Tabellenverzeichnis

2.1. Standardfehler und Signifikanzniveaus der multivariaten Regression	19
2.2. Bestimmtheitsmaße und Standardfehler der multivariaten Regression . . .	20
2.3. Winkelwerte eines fiktiven Individuums	22
3.1. Möglichkeiten der Klassifizierung	29
4.1. Standardfehler und Signifikanzniveaus für die bivariate Regression	36
4.2. Bestimmtheitsmaße und Standardfehler der bivariaten Regression	36
5.1. Vergleich der multivariaten mit der bivariaten Regression	48
5.2. Klassifikation der Individuen im jeweiligen Ansatz	49
5.3. Klassifikation: Harmoniebox mit 2 Standardfehlern und MH mit $\lambda_{\hat{\sigma}}^{(j)}(2)$.	49
A.1. Korrelationen innerhalb der Normgruppe	62